

## Peptide information for frame 1

ORF from 202 bp to 876 bp; peptide length: 225  
Category: similarity to known protein

1 MNQPCNSMEP RVMDDMLKL AVGDQGPQEE AGQLAKQEGI LFKDVLSLQL  
51 DFRNLRIDN LWQFENLRKL QLDNNIEKI EGLNLAHLV WLDLSFNNE  
101 TIEGLDTLVN LEDLSLFNNR ISKIDSLDAL VKLQVLSLGN NRIDNMMNII  
151 YLRRFKCLRT LSLSRNPISE AEDYKMFICA YLPDLMYLDY RRIDDHTASV  
201 SLSVSQPCET DSSSPQVSWK RGIEE

## BLASTP hits

Entry S68209 from database PIR:  
sds22 protein homolog - human >TREMBL:HSSDS22MR\_1 gene: "sds22";  
product: "yeast sds22 homolog"; H.sapiens sds22-like mRNA  
Score = 234, P = 1.2e-19, identities = 61/143, positives = 93/143

Entry A38439 from database PIR:  
suppressor protein sds22(+) - fission yeast (Schizosaccharomyces pombe)  
>SWISSPROT:SD22\_SCHPO PROTEIN PHOSPHATASES PP1 REGULATORY SUBUNIT  
SDS22. >TREMBL:SPAC4A8\_12 gene: "sds22"; product: "phosphatases pp1  
regulatory subunit"; S.pombe chromosome I cosmid c4A8.  
Score = 208, P = 5.6e-17, identities = 52/127, positives = 71/127

Entry S43988 from database PIR:  
protein suppressor sds22 - fission yeast (Schizosaccharomyces pombe)  
>SWISSPROT:SD22\_SCHPO PROTEIN PHOSPHATASES PP1 REGULATORY SUBUNIT  
SDS22. >TREMBL:SPAC4A8\_12 gene: "sds22"; product: "phosphatases pp1  
regulatory subunit"; S.pombe chromosome I cosmid c4A8.  
Score = 208, P = 8.5e-17, identities = 52/127, positives = 71/127

Entry CEK10D2\_5 from database TREMBL:  
gene: "K10D2.1"; Caenorhabditis elegans cosmid K10D2.  
Score = 214, P = 3.6e-16, identities = 50/125, positives = 75/125

## Alert BLASTP hits for DKFZphut1\_20ml1, frame 1

No Alert BLASTP hits found

## Pedant information for DKFZphut1\_20ml1, frame 1

## Report for DKFZphut1\_20ml1.1

[LENGTH]	225
[MW]	25955.87
[pI]	4.63
[HOMOL]	PIR:S68209 sds22 protein homolog - human 1e-18
[FUNCAT]	03.22 cell cycle control and mitosis [S. cerevisiae, YKL193c] 2e-11
[FUNCAT]	30.10 nuclear organization [S. cerevisiae, YKL193c] 2e-11
[FUNCAT]	06.07 protein modification (glycosylation, acylation, myristylation, palmitoylation, farnesylation and processing) [S. cerevisiae, YKL193c] 2e-11
[FUNCAT]	30.05 organization of centrosome [S. cerevisiae, YOR373w] 2e-06
[FUNCAT]	01.03.10 metabolism of cyclic and unusual nucleotides [S. cerevisiae, YJL005w] 3e-05
[FUNCAT]	03.10 sporulation and germination [S. cerevisiae, YJL005w] 3e-05
[FUNCAT]	30.02 organization of plasma membrane [S. cerevisiae, YJL005w] 3e-05
[FUNCAT]	10.04.03 second messenger formation [S. cerevisiae, YJL005w] 3e-05
[FUNCAT]	04.07 rna transport [S. cerevisiae, YPL169c] 9e-04
[FUNCAT]	04.05.01.04 transcriptional control [S. cerevisiae, YCR065w] 9e-04
[EC]	4.6.1.1 Adenylate cyclase 2e-06
[PIRKW]	nucleus 5e-16
[PIRKW]	duplication 2e-06
[PIRKW]	tandem repeat 2e-06
[PIRKW]	cAMP biosynthesis 2e-06
[PIRKW]	glycoprotein 2e-06
[PIRKW]	phosphorus-oxygen lyase 2e-06
[SUPFAM]	leucine-rich alpha-2-glycoprotein repeat homology 5e-16
[SUPFAM]	fibromodulin 3e-07
[SUPFAM]	yeast adenylate cyclase catalytic domain homology 2e-06
[SUPFAM]	yeast adenylate cyclase 2e-06
[PROSITE]	CK2_PHOSPHO_SITE 2
[PROSITE]	PKC_PHOSPHO_SITE 1

[KW] All\_Alpha

```

SEQ  MNQPCNSMEPRVMDDDMLKLAVGDQGPQEEAGQLAKQEGILFKDVLSLQLDFRNILRIDN
PRD  cccccccccccccchhhhhccccccccchhhhhhhhhhhchhhhhhhhhccccccccccccc

SEQ  LWQFENLRKLQLDNNIEKIEGLENLHLVWLDLSFNNIETIEGLDTLVNLEDLSLFNNR
PRD  hhhhhhhhhhhccccccccccccchhhhhhhccccccccccccchhhhhhhhhcccccc

SEQ  ISKIDSLDALVKLQVLSLGNNRIDNMMNIIYLRREFKCLRTLSLSRNPISEAEDYKMFICA
PRD  cccchhhhhhhhhhhccccccccccccccccchhhhhhhccccccccccccchhhhhhhhh

SEQ  YLPDLMYLDYRRIDDHTASVLSVSPQPCETDSSSPQVSWKRGIEE
PRD  hhccccccccccccchhhhhhhcccccccccccccccccccccc

```

Prosite for DKFZphutel\_20m11.1

PS00005	218->221	PKC_PHOSPHO_SITE	PDOC00005
PS00006	122->126	CK2_PHOSPHO_SITE	PDOC00006
PS00006	169->173	CK2_PHOSPHO_SITE	PDOC00006

(No Pfam data available for DKFZphutel\_20m11.1)

DKFZphute1\_20m24

group: metabolism

DKFZphute1\_20m24 encodes a novel 611 amino acid protein with similarity to a hypothetical *C.elegans* protein and to yeast Alg9 protein.

This protein is a putative mannosyl transferase that is involved in the assembly of the core oligosaccharide Glc3Man9GlcNAc2.

The new protein can find application in modulation of glycosylation of proteins and as a new enzyme for biotechnologic production processes.

strong similarity to *S.cerevisiae* Alg9p

complete cDNA, complete cds, potential start at Bp 23, few EST hits  
Alg9 is involved in the assembly of the core oligosaccharide  
Glc3Man9GlcNAc2

HSAC381 corresponding genomic DNA (2 exons)

HSB8954 corresponding genomic DNA (1 exon)

Sequenced by AGOWA

Locus: /map="11"

Insert length: 1986 bp

Poly A stretch at pos. 1966, polyadenylation signal at pos. 1949

```
1  TTCTTTTTC CCCAGGCTTG CCATGGCTAG TCGAGGGGCT CGGCAGCGCC
51  TGAAGGGCAG CGGGGCCAGC AGTGGGGATA CGGCCCGGCG TCGGACAAG
101 CTGCGGGAGC TGCTGGGCGAG CCGAGAGGCG GGCGGCGCGG AGCACCGGAC
151 CGAGTTATCT GGGAAACAAAG CAGGACAAGT CTGGGCACCT GAAGGATCTA
201 CTGCTTTCAA GTGCTGCTT TCAGCAAGGT TATGTGCTGC TCTCCTGAGC
251 AACATCTCTG ACTGTGATGA AACATTCAAC TACTGGGAGC CAACACACTA
301 CCTCATCTAT GGGGAAGGGT TTCAGACTTG GGAATATTCC CCAGCATATG
351 CCATTCGGTC CTATGCTTAC CTGTTGCTTC ATGCCTGGCC AGCTGCATT
401 CATGCAAGAA TTCTACAAAC TAATAAGATT CTTGTGTTTT ACTTTTTCG
451 ATGTCTTCTG GCTTTTGTGA GCTGTATTG TGAACCTTAC TTTTACAAGG
501 CTGTGTGCAA GAAGTTTGGG TTGCACGTGA GTCGAATGAT GCTAGCCTTC
551 TTGGTCTCTA GCACTGGCAT GTTTTGCTCA TCATCAGCAT TCCTTCCTAG
601 TAGCTTCTGT ATGTACACTA CGTTGATAGC CATGACTGGA TGGTATATGG
651 ACAAGACTTC CATTGCTGTG CTGGGAGTAG CAGCTGGGGC TATCTTAGGC
701 TGGCCATTCA GTGCAGCTCT TGGTTTACCC ATTGCCTTTG ATTTGCTGGT
751 CATGAAACAC AGGTGGAAGA GTTCTTTTCA TTGGTCGCTG ATGGCCCTCA
801 TACTATTCTT GGTGCCTGTG GTGGTCATG ACAGCTACTA TTATGGGAAG
851 TTGGTGATTG CACCACTCAA CATTGTTTTG TATAATGTCT TTAATCTCA
901 TGGACCTGAT CTTTATGGTA CAGAACCTCG GTATTCTAT TTAATTAATG
951 GATTTCGAA TTTCAATGTA GCCTTTGCTT TGGCTCTCCT AGTCCTACCA
1001 CTGACTTCTC TTATGGAATA CCTGCTGCAG AGATTTCATG TTCAGAATTT
1051 AGGCCACCCG TATTGGCTTA CCTTGGCTCC AATGTATATT TGGTTTATAA
1101 TTTCTTTCAT CCAGCCTCAC AAAGAGGAGA GATTTCCTTT CCCTGTGTAT
1151 CCACCTTATG GTCTCTGTGG CGCTGTGGCT CTCTCTGCAC TTCAGAAATG
1201 TTACCACTTT GTGTTTCAAC GATATCGCCT GGAGCACTAT ACTGTGACAT
1251 CGAATTGGCT GGCATTAGGA ACTGTCTTCC TGTTTGGGCT CTTGTCATT
1301 TCTCGCTCTG TGGCACTGTT CAGAGGATAT CACGGGCCCG TTGATTTGTA
1351 TCCAGAATTT TACCGAATTG CTACAGACCC AACCATCCAC ACTGTCCAG
1401 AAGGCAGACC TGTGAATGTC TGTGTGGGAA AAGAGTGGA TCGATTTCCC
1451 AGCAGCTTCC TTCTTCCTGA CAATTGGCAG CTTCAGTTCA TTCCATCAGA
1501 GTTCAGAGGT CAGTTACCAA AACCTTTTGC AGAAGGACCT CTGGCCACCC
1551 GGATTGTTCC TACTGACATG AATGACCAGA ATCTAGAAGA GCCATCCAGA
1601 TATATTGATA TCAGTAAATG CCATTATTTA GTGGATTGG ACACCATGAG
1651 AGAAACACCC CGGGAGCCAA AATATTCATC CAATAAAGAA GAATGGATCA
1701 GCTTGGCCTA TAGACCATTC CTTGATGCTT CTAGATCTTC AAAGCTGCTG
1751 CGGGCATTCT ATGTCCCTT CCTGTGAGAT CAGTATACAG TGTACGTAAA
1801 CTACACCATC CTCAAACCCC GGAAGGCAAA GCAATCAGG AAGAAAAGTG
1851 GAGGTTAGCA ACACACCTGT GGCCCCAAG GACAACCATC TTGTTAACTA
1901 TTGATTCCAG TGACCTGACT CCCTGCAAGT CATCGCCTGT AACATTTGTA
1951 ATAAAGGTCT TCTGACATGA AAAAAAAAAA AAAAAA
```

## BLAST Results

Entry HSAC381 from database EMBL:  
Homo sapiens chromosome 11 pac pDJ159o1, complete sequence.  
Length = 42,771

Entry HSB8954 from database EMBL:

cSRL-50A3-u cSRL flow sorted Chromosome 11 specific cosmid Homo sapiens genomic clone cSRL-50A3.  
Length = 601

# Medline entries

96293493:  
Stepwise assembly of the lipid-linked oligosaccharide in the endoplasmic reticulum of *Saccharomyces cerevisiae*: identification of the ALG9 gene encoding a putative mannosyl transferase.

# Peptide information for frame 2

ORF from 23 bp to 1855 bp; peptide length: 611  
Category: strong similarity to known protein

```

1 MASRGARQRL KGSGASSGDT APAADKLREL LGSREAGGAE HRTLSGNKA
51 GQVWAPEGST AFKCLLSARL CAALLSNISD CDETFNYWEP THYLIYGEFG
101 QTWEYSPAYA IRSYAYLLH AWPAAFHARI LQTNKILVEY FLRCLLAFVS
151 CICELYFYKA VCKKFGHLVS RMMLAFLVLS TGMFCSSSAF LPSSFCMYTT
201 LIAMTGWYMD KTSIAVLGVA AGAILGWPF S AALGLPIAFD LLVMKHRWKS
251 FFHWSLMALI LFLVPVVVID SYYYGKLVIA PLNIVLYNVF TPHGPDLYGT
301 EPWYFYLING FLNFNVAFAL ALLVLPLTSL MEYLLQRFHV QNLGHPYWLT
351 LAPMYIWFII FFIQPHKEER FLFPVYPLIC LCGAVALSAL QKCYHFVFQR
401 YRLEHYTVTS NWLALGTVFL FGLLSFSRSV ALFRGYHGPL DLYPEFYRIA
451 TDPTIHTVPE GRPVNVCVGK EWYRFPSSFL LPDNWQLQFI PSEFRGQLPK
501 PFAEGPLATR IVPTDMNDQN LEEPSRYIDI SKCHYLVOLD TMRETPREP K
551 YSSNKEEWS LAYRPFLDAS RSSKLLRAFY VPFLSDQYTV YVNYTILKPR
601 KAKQIRKKSG G

```

# BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphutel\_20m24, frame 2

SWISSPROT:YTH3\_CAEEL HYPOTHETICAL 75.5 KD PROTEIN C14A4.3 IN CHROMOSOME II., N = 1, Score = 957, P = 2.7e-96

PIR:S63177 mannosyl transferase (EC 2.4.1.-) - yeast (*Saccharomyces cerevisiae*), N = 1, Score = 533, P = 2.3e-51

SWISSPROT:YTH3\_CAEEL HYPOTHETICAL 75.5 KD PROTEIN C14A4.3 IN CHROMOSOME II., N = 1, Score = 957, P = 2.7e-96

PIR:S63177 mannosyl transferase (EC 2.4.1.-) - yeast (*Saccharomyces cerevisiae*), N = 1, Score = 533, P = 2.3e-51

>SWISSPROT:YTH3\_CAEEL HYPOTHETICAL 75.5 KD PROTEIN C14A4.3 IN CHROMOSOME II.

Length = 653

# HSPs:

Score = 957 (143.6 bits), Expect = 2.7e-96, P = 2.7e-96  
Identities = 206/514 (40%), Positives = 296/514 (57%)

```

Query:  48 NKAGQVWAPEGSTAFKCLLSARLCAALLSNISDCDETFNYWEPH YLIYGEFGQTWEYSP 107
          N   W   + FK LLS R+  A+  I+DCDE +NYWEP H  +YGEFGQTWEYSP
Sbjct:  43 NNPDNDWPFSFGSVFKMLLSIRISGAIWGIINDCDEVYNYWEPLHLFLYGEFGQTWEYSP 102

Query:  108 AYAIRSYAYLLHAWPAAFHARILQTNKILVEYFLRCLLAFVSCICELYFYKAVCKKFG 167
          YAIRSY Y+ LH  PA+  A+  KI+VF +R  +  + E Y + A+CKK  +
Sbjct:  103 VYAIRSYFYIYLHYIPASLFANLFGDTRIVVFTLIRLTIGLFCLLGEYAFDAICKKINI 162

Query:  168 HVSRMMLAFLVLSTGMFCSSSAFLPSSFCMYTTLIAMTGWYMDKTSIAVLGVAAGAILGW 227
          R  + F + S+GMF +S+AF+PSSFCM T  +  +  +  +  + VA  ++GW
Sbjct:  163 ATGRFFILFSIFSSGMFLASTAFVPSSFCMAITFYILGAYLNENWTAGIFCVAFSTMVGW 222

Query:  228 PFSAALGLPIAFDLLVMKHRWKSFFHWSLMALILFLVPVVVIDSYYYGKLVIAPLNIVLY 287

```



Sbjct: 223 PFSAVLGLPIVADMLLLKGLRIRFILTSLVIGLCIGGVQVITDSHYFGKTVLAPLNIFLY 282

Query: 288 NVFTPHGPDLYGTEPWYFYLINGFLNFNVAFAALALLVPLTSLMEYLLQRFHVQNLGHPY 347  
 NV + GP LYG EP FY+ N F N+N+ A PL+ + Y + + Q+

Sbjct: 283 NVVSGPGPSLYGEEPLSFYIKNLFNNWNIVIFAAPFGFPLS--LAYFTKVWMSQDRNVAL 340

Query: 348 WLTAPMYI-----WFIIFFIQPHKEERFLFPVYPLICGAVALSALQKCYHFVFOR 400  
 + AP+ + W +IF Q HKEERFLFP+YP I A+AL A + ++

Sbjct: 341 YQRFAPILLAVTTAAWLLIFGSQAHEERFLFPYIPFIAFFAALALDATNR---LCLKK 397

Query: 401 YRLEHYTVTSNWLALGTVFLFGLLSFSRSVALFRGYHGPLDLYPEFYRIATDEPTIHTVPE 460  
 ++ N L++ + F +LS SR+ ++ Y +++Y T+ T +

Sbjct: 398 LGMD-----NILSILFILCFAILSASRTYSIHNNYGSVHEIYRSLNAELTNRT-NFKNF 450

Query: 461 GRPVNVCGKWEYRFPSSFLPDNW-----QLQFIPSEFRGQLPKPFAEGPL---ATRI 511  
 P+ VCVGKEW+RFPSSF +P +++FI SEFRG LKPKF + TR

Sbjct: 451 HDPIRVCVGKEWHRFPSFFIPQTVSDGKKVEMRFIQSEFRGLLPKPFLLKSDKLVEVTRH 510

Query: 512 VPTDMNDQNLEEPSRYIDISKCHYLVDLDTMRETREPKEYSSNKEEW 558  
 +PT+MN+ N EE SRY+D+ C Y+VD+D M ++ REP + ++ +

Sbjct: 511 IPTEMNNLNQEEISRYVDLSDCDYVVDVD-MPQSDREPDRKMRQNY 556

Pedant information for DKFZphutel\_20m24, frame 2

# Report for DKFZphutel\_20m24.2

[LENGTH] 611  
 [MW] 69863.78  
 [pI] 8.91  
 [HOMOL] SWISSPROT:YTH3\_CAEEL HYPOTHETICAL 75.5 KD PROTEIN C14A4.3 IN CHROMOSOME II. 2e-93  
 [FUNCAT] 09.01 biogenesis of cell wall [S. cerevisiae, YNL219c] 4e-69  
 [FUNCAT] 01.06.01 lipid, fatty-acid and sterol biosynthesis [S. cerevisiae, YNL219c] 4e-69  
 [FUNCAT] 01.05.01 carbohydrate utilization [S. cerevisiae, YNL219c] 4e-69  
 [PIRKW] glycosyltransferase 9e-68  
 [PIRKW] transmembrane protein 9e-68  
 [PIRKW] hexosyltransferase 9e-68  
 [PROSITE] MYRISTYL 9  
 [PROSITE] CAMP\_PHOSPHO\_SITE 1  
 [PROSITE] CK2\_PHOSPHO\_SITE 7  
 [PROSITE] PKC\_PHOSPHO\_SITE 6  
 [PROSITE] ASN\_GLYCOSYLATION 2  
 [KW] TRANSMEMBRANE 7  
 [KW] LOW\_COMPLEXITY 6.71 %

SEQ MASRGARQRLKSGASSGDTAPAADKLELLGSREAGGAHRTLSGNKAGQVWAPEGST  
 SEG .....  
 PRD ccchhhhhhhccccccccccccchhhhhhhhhccccccccccccccccccccccccch  
 MEM .....MMMMMM

SEQ AFKCLLSARLCAALLSNISDCDETFFNYWEPHYLIYGEFGTWEYSPAYAIRSYAYLLLH  
 SEG .....xxxxxxxxxxxxxxxxx.....  
 PRD hhhhhhhhhhhhhhhhhhhhhccch  
 MEM MM

SEQ AWPAAFHARILQTNKILVFYFLRCLLAFVSCICELYFYKAVCKKFGHLVSRMMLAFLVLS  
 SEG .....  
 PRD cchhhcc  
 MEM MM

SEQ TGMFCSSSAFLPSSFCMYTTLIAMTGWYMDKTSIAVLGVAAGAILGWPFSAALGLPIAFD  
 SEG .....xxxxxxxxxxxxxxxxxxxxx.....  
 PRD cccccccccccccchhhhhhhhhhhhhccccccccccccccccccccccccccccccccccccch  
 MEM .....MMM

SEQ LLVMKRWKSFHWSLMALILFLVPVVVIDSYYYGKLVIAPLNIVLYNVFTPHGPDLYGT  
 SEG .....  
 PRD hhhcc  
 MEM MMMMMMM.MM

SEQ EPWYFYLINGFLNFNVAFAALALLVPLTSLMEYLLQRFHVQNLGHPYWLTLAPMYIWFII  
 SEG .....xxxxxxxxxxxxxxxxxxxxx.....  
 PRD cccccccccccccchhh  
 MEM .....MMM

Prosite for DKFZphute1\_20m24.2

(No Pfam data available for DKFZphute1\_20m24.2)

DKFZphut1\_21d15

group: uterus derived

DKFZphut1\_21d15 encodes a novel 191 amino acid protein without similarity to known proteins.

No informative BLAST results; No predictive prosite, pfam or SCOP motife.

The new protein can find application in studying the expression profile of testis-specific genes.

unknown

Sequenced by MediGenomix

Locus: /chromosome="3"

Insert length: 5292 bp

Poly A stretch at pos. 5273, polyadenylation signal at pos. 5252

```
1 CTCCCACTAG TGTATGCCTT AATGGTGCCG CTCTTGTCGG CGTCTACGCT
51 TGGGACCTTG GCTTCTGACT TGGAGAGTGT ACAGCTCTGC CCGACGGCAA
101 CCCAGCTTGG GAAGAGAAGC CCCAGCGTGG GCTGGGGCTC AAGGCGCAGG
151 AAGGCCGAGC CCGGCGCGGA CGCAGGCGGC TCCGGGCGGG CTCAGCACCC
201 CCAGGCACCG TCTCCTAGTG ACCGCGGCGC TCGCGGGCCT GCGGCGCGTT
251 GTCCGGGCGA CTGCGCAGCG CGGGCACCCC CGCGGCCCTT CCCCTGGGCG
301 CGCGCGCGAC CTGGGTGCCA TGGCGGCAGC GCGCGTGACA GGCCAGCGGC
351 CTGAGACCGC GCGGCGCGAG GAGGCTCGA GGCCGCAGTG GCGCGCGCCA
401 GACCACTGCC AGGCTCAGGC GCGGCGCGGG CTGGGCGACG GCGAGGACGC
451 ACCGGTGCGT CCGCTGTGCA AGCCCCGCGG CATCTGCTCG CGCGCCTACT
501 TCCTGGTGCT GATGGTGTTT GTGCACCTGT ACCTGGGTAA CGTGCTGGCG
551 CTGCTGCTCT TCGTGCACTA CAGCAACGGC GACGAAAGCA GCGATCCCGG
601 GCCCCAACAC CGTGCCCAAG GCCCGGGGCG CGAGGCCACC TTAGGTCCCC
651 TCACCCGGCT GGAGGGCATC AAGGTGAGGA CCTCCCTGCC CCGCCGCGCT
701 CCAGGCCCTG CACGGCTGAG CCCGAGAGGA CCGGCGCTCA GCGCGGGTCC
751 CCACGCTGCC CCGGCGCTG CTCTGCGTCG GTCCCGCGCG CTCCCACTCA
801 CTCGCTGCTT GTCGCTCTCC GGGCGGGGCG GACTTGGCCC TTTTGGGCA
851 GCGCGGTCTG GCGCCCCAGC TGCCCGCTGT GCGCCTTTTC CTTAGGTGGG
901 GCACGAGCGT AAGGTCCAGC TGGTCACCGA CAGGGATCAC TTCATCCGAA
951 CCCTCAGCCT CAAGCCGCTG CTCTTCGAAA TCCCCGGCTT CCTGACTGAT
1001 GAAGAGTGTC GGCTCATCAT CCATCTGGCG CAGATGAAGG GGTACAGCG
1051 CAGCCAGATC CTGCTACTG AAGAGTATGA AGAGGCAATG AGCACTATGC
1101 AGGTACAGCA GCTGGACCTC TTCCGGCTGC TGGACCAGAA CCGTGATGGG
1151 CACCTTCAGC TCCGTGAGGT TCTGGCCAG ACTCGCCTGG GAAATGGATG
1201 GTGGATGACT CCAGAGAGCA TTCAGGAGAT GTACGCCGCG ATCAAGGCTG
1251 ACCCTGATGG TGACGGTGAG CTCACACCTC TGCACAGTCC TATCCCGTG
1301 AGCCTCCTGC CCACTCCAG GTGCACAATT TTGAAAACCT GGGCCCTTCC
1351 CCCACAGCCA GGCAGCCTCT CTGCACCCCT TTATAGTGGC CAGAGATGGG
1401 GAGGTGAAGA TCCAGCCTTG CTTTTTACCC CTGGGAAGTA GGCAGGCAGC
1451 CAGGCCCCCC GTTCCCTTG GTGATGGTCT CGAGGGCAGT TCTTGAGAC
1501 CCTTTGATA ACATCAGGCA GAGTTGAGAG CCTGGGGACA GGAAGTAGGG
1551 CTGCTAGTTG GCAGAGAACA GAGTGGGTGG AGCAGGAGCA AGGCGACAGT
1601 GAGGCCAGCT AGAGCTTGGC TGTTTACCTT GCTCCATCCA TCTCTCCAGC
1651 CAGACACGAG GTCCACCCCA GCAGACAGCT TCCCTGGTCT AAGTAGGCTC
1701 TCCCTTGCCT TCCTCTTGTC CACCTGGAGT CATGCCGAAG CGCCTAAAAT
1751 GGTAGTGCTG CTACCTGTGC TAACTGCTGG GGAGGGGTGG GCAGGGAAGC
1801 TGTCATGCAA GTGGTGCCCC CTCTGGTAAT AACTCTCAGG AGGTTTCTGA
1851 GGTGTGGTCA TCACCTCAT GCCCAAATTC TGGACCAAGA GAGGAAGATA
1901 CAGCAGTTAG AAAGGACTTG GAACAGTGGC TTTGCGGCTG GTGAACCAGA
1951 GTGAAGAATC TGGCCGTGAC CTGGCTGCCA CACTGCTATA GGCCCCAGAA
2001 CAGAGGTGGT GACAGTCTCA CAGCCCTTGA ATGTCCCCCA CCCTCAGAGG
2051 AATCTGGGCC AAAGAGTGGA AGGTGATGTC CTTGGGTGAG CCAGAATAAC
2101 ATGGAGCAAA GATACCAACT ACTCTCCAG AACCCCAAGA GGGTAGAACC
2151 CCTGCTTAAT GGTTTGAGCA GGGACAGTGG AGAATGTTCT CATGAGAGGG
2201 GGTGGCCTGA CTTTCGTGTC TAAAGTGGCT GGTAAACGAG TAGGCAGGGC
2251 TGGCGAAGTA GGTTCACCC AGGATGAAC CTGGGGTCAT GAGGAACTCC
2301 CCGGGGGCTG GCCCTGCTTG CACCTGGCG TATGTATGTA AGGCCCTGGA
2351 TGAGGCCAG CACTGCCTGC TCTCTCTCA CCCTCCACAG GCCGGAGAGT
2401 GGCCACCACT CTATATAGCC AGGCTGGAAG GCCAGGGTCC TGGCCATATG
2451 GCTCAAGCTT CCTTTGGAGA ACCTTCTCTG GCCACTCTAA TAGGGGTGG
2501 GCCTCTTTCT TCTTAGGGCC AAATTAGGGC TTAAGTCTAG AAAAGGAAGT
2551 GCTCTGGGTC TTCCTGTAAG GCCTGATGTG ACAGAAACCA GGTTCATCTG
2601 ACCCAAAAGT CCAGGTGGGG GACAAGTGTG CAAGGCCCTT CAGTGCCTGA
2651 GGTCAAGGGC TGCTGCTGCC TTTGGGGTAG GTAGGGAAGT GCAGCCTGCC
2701 ACTGTTGCCCT CCAATATAGG GCTTGGTGGG CATTGATGGT GGGTGCCCTG
2751 TGCAGGAGTG CTGAGTCTGC AGGAGTTCTC CAACATGGAC CTTGGGGACT
2801 TCCACAAGTA CATGAGGAGC CACAAGGCAG AGTCCAGTGA GCTGGTGCGG
```

```

2851 AACAGCCACC ATACCTGGCT CTACCAGGGT GAGGGTGCCC ACCACATCAT
2901 GCGTGCCATC CGCCAGAGGT GAGCAGCTGA AGCTGTTCTC ACTGGAGCAG
2951 GGGGAGAAGA CTGGGCAGGG CCTCCACAGA AGCTCTTGTC TGGGGCCAAG
3001 AGGACAGAAT GGATTAACCC ATTTGGGATT AAGTTCCATT TGTTAGACCA
3051 GGATTGGGAC CCACTGAAAG ACAGGCAATT AACAAAGGCA AATTAGCCCT
3101 CCTTGCAGGC ACACAATGGG CAACTGGGGT TAGATAGAGA TTGAGCACTT
3151 CTTTCTGATT AGATAAATGA CCTCTTATCT TTGACCCCTT ATCTGACCCC
3201 GTCACAGCAG GAAAAGGGTT TTTAAATAAA CAACTTTCTT CCAGGGAGGA
3251 GGACCTCAGG ACTCCCGGCC CCCTTTATTT AGTGGAAATG TCAACATTTC
3301 CACATAGCAG GTGCTCTGTG CTTTGGCATC TGAGGGAGAA GGATCATCAT
3351 GAGTAACCCC CTCCTGCTCT TACAGGGCCA GTCTGAGATG GCTTAAGGGA
3401 CTTCCAGGGG AGGTGGGTAG GGGCAAAGCT TGTGGCAGGC CTAGGGTCCA
3451 CCTTGGCCAG CTCCTTCAGA TCACCACCTT GCCTGGGGCT GCCCAGCCAA
3501 ATGCCTCTGT CCCACCAGGG TGCTGCGCCT CACTCGCCTG TCGCCTGAGA
3551 TCGTGGAGCT CAGCGAGCCG CTGCAGGTTG TTCGATATGG TGAGGGGGGC
3601 CACTACCATG CCCACGTGGA CAGTGGGCCCT GTGTACCCAG AGACCATCTG
3651 CTCCCATAGC AAGCTGGTAG CCAACGAGTC TGTACCCTTC GAGACCTCCT
3701 GCCGGCAAGT ATCTCCCAAC TGGGGGCTGC CTTCAATCCT CAGACCAGGA
3751 ACACCCATGA CACAGGCACA GCCCTGCACT GTGGGCGTGC CCCTTGGCAT
3801 GGGGCCAGGA GATCACTGGG TTATCCCGGT TAGTGATGCC CTCACCTCTC
3851 CCACCAAGTT GTTTACCCAA TGGCTGGAAA GGGGTGGGTA CTGGTCATCG
3901 TGACCACTGG AGTCAACACA GACTGATGTA CCCACAGACA CCAAACTTG
3951 CCCCTGAGT TCTGAAGCAA GGGGCAAGGC TGGGCCCCTA GCTTGCTCTG
4001 CCCATTCCTC CAGGTGTTGA TCTTGATTCC ACTTAGAGAA GCTGAAGCTG
4051 TGCCCTCCCTC CCCTGTCAAG CCAGTCTTTT CCTCTTCAGG TGGGTGTTCT
4101 GGCCAGGCC CTTCCCATCC CCAAGGAGCC CTTACGCGG CCCTGTTGCT
4151 TCTGCTAGCC TACCTTTCCC TGCCAGGCC TTGCTCAGGG CCATGGCATT
4201 TAACTAAGTG CACCTGTGAT CTTGGCCAAA AAACCATTGC AACTCACAGT
4251 AAGAGACTGG GTTTCGGGGA AGGAGGGGCT AGGGACATTT TGGCACTGGC
4301 CTGCCCTATT GTCTCCATC CTAGTCTGTC CTGGTCCCTG GCAACAGGAA
4351 CCTGGGCAGC TTATCCTGCC CACAGGTAAG CCCCTGGGAG CATCCACAAC
4401 TGGGGACCTG CTCAGTGCCC CCCCTGCCCT ACAGCTACAT GACAGTGCTG
4451 TTTTATTGTA ACAACGTCAC TGGTGGGGGC GAGACTGTTT TCCCTGTAGC
4501 AGATAACAGA ACCTACGATG AAATGGTAAG GGTCAACTGG GCTATTACTC
4551 TTGTGGGCTG GCAGGGGCTT AGACAAGTGA AGTACACACC TCTCCAGGTC
4601 TAAGGATGTG GGCCCAAATT ATTCTTGGG CATATCTGGT TGGTTTCCCT
4651 TTGGTCACCC TTGGCTGGCC TGGCCATAGA GTGGGGACAG GTTGAACACC
4701 CCACCACCCT GCTGCCACA GAGTCTGATT CAGGATGACG TGGACCTCCG
4751 TGACACACGG AGGCACTGTG ACAAGGGAAC CCTGCGTGTC AAGCCCAAC
4801 AGGGCACAGC AGTCTTCTGG TACAACTACC TGCCTGATGG GCAAGGTTGG
4851 GTGGGTGACG TAGACGACTA CTCGCTGCAC GGGGCTGCC TGGTCACGCG
4901 CGGCACCAAG TGATTGCCA ACAACTGGAT TAATGTGGAC CCCAGCCGAG
4951 CGCGGCAAGT GCTGTTCCAA CAGGAGATGG CCCGCCCTGC CCGAGAAGGG
5001 GGCACCGACT CACAGCCCGA GTGGGCTCTG GACCGGGCCT ACCGCGATGC
5051 GCGCGTGGAA CTCTGAGGGA AGAGTTAGCC CCGGTTCCTA GCCGCGGGTC
5101 GCCAGTTGCC CAAGATCAGG GGTCCGGCTG TCCTTCTGTC CTGCTGCAGA
5151 CTAAGGTCTT GGCCAATGTC TTGCCCCACC CCGCCAGCCG CGATACGGCG
5201 CAGTTCCTAT ATTCAATGTA TTTATTGTGT ACTGACTCCA TCTGCCCCGT
5251 CAAATAAAAA ACCACAAGGT TCGAAAAAAA AAAAAAAA GG

```

## BLAST Results

Entry HSU64252 from database EMBL:  
Human STS sequence NOTI-225.  
Score = 959, P = 1.2e-36, identities = 195/199

## Medline entries

No Medline entry

## Peptide information for frame 1

ORF from the beginning to 351 bp; peptide length: 118  
Category: questionable ORF  
Classification: no clue

```

1 LPLVYALMVP LLSASTLGTL ASDLESVQLC PTATOLGKRS PSVWGSRRR
51 KAEPGADAGG SGRAQHPQAP SPSDRGARGP GGRCPGDCAA RAPPRPLPWA
101 RARPGCHGGS GGDPRPAA

```

## BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphut1\_21d15, frame 1

No Alert BLASTP hits found

Peptide information for frame 2

ORF from 320 bp to 892 bp; peptide length: 191  
Category: putative protein  
Classification: no clue

1 MAAAVTQGR PETAAEEAS RPQWAPPDHC QAQAAAGLGD GEDAPVRPLC  
51 KPRGICSRAY FLVLMVFVHL YLGNVLALLL FVHYSNGDES SDPGPQHRAQ  
101 GPGPEPTLGP LTRLEGIKVR TSLPRRAPGP ARLSPRGPAL SPGPHAAPGA  
151 ALRRSRALPL TRLLSLSGPG RLGPFWAARS GAPAAARCAPF P

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphut1\_21d15, frame 2

PIR:EDBE75 immediate-early protein IE175 - human herpesvirus 1, N = 2,  
Score = 106, P = 0.0067

>PIR:EDBE75 immediate-early protein IE175 - human herpesvirus 1  
Length = 1,298

HSPs:

Score = 106 (15.9 bits), Expect = 6.7e-03, Sum P(2) = 6.7e-03  
Identities = 36/103 (34%), Positives = 44/103 (42%)

Query: 87 GDESSDPGPQHRAQGGPGPEPTLGLPLTRLEGIKVRTSLPRRA-PGPARLS-PRGPALSPGP 144  
G + PGP G GP P P T+ G S R P PA S P GP +P  
Sbjct: 726 GRKRKSPGPAPPPGGGGPRP---PKTKKSGADAPGSDARAPLPAPAPPSTPPGPEPAPAQ 782

Query: 145 HAAPGAALRRSRALPLT-RLLSLSGPGRLGPFWAARS GAPAAARCAP 189  
AAP AA ++R P+ GP LG W + P+ AP  
Sbjct: 783 PAAPRAAAQARPRPVAVSRRAEGPDPLGG-WRRQPPGPSHTAAP 827

Score = 40 (6.0 bits), Expect = 6.7e-03, Sum P(2) = 6.7e-03  
Identities = 8/21 (38%), Positives = 9/21 (42%)

Query: 28 DHCQAQAAAGLGDGEDAPVRP 48  
DH + A G G AP P  
Sbjct: 212 DHAREARAVGRGPSSAAPAAP 232

Pedant information for DKFZphut1\_21d15, frame 1

Report for DKFZphut1\_21d15.1

[LENGTH] 117  
[MW] 11797.32  
[pI] 10.68  
[KW] Irregular  
[KW] SIGNAL PEPTIDE 22  
[KW] LOW\_COMPLEXITY 38.46 %

SEQ LPLVYALMVPLLSASTLGTLASDLESVQLCPTATQLGKRSPSVGWSRRRKAEPGADAGG  
SEG .....xx  
PRD cccccccccccccccccccccchhhhhhhhhcccccccccccccccccccccccccccccccccccc

SEQ SGRAQHPQAPSPSDRGARGPGGRCFPGDCAARAPPRPLPWARARPGCHGGSGGDRPAA  
SEG .....xx  
PRD ccc

(No Prosite data available for DKFZphut1\_21d15.1)

(No Pfam data available for DKFZphut1\_21d15.1)

Pedant information for DKFZphut1\_21d15, frame 2

Report for DKFZphut1\_21d15.2

```
{LENGTH}      191
{MW}           19916.88
{pI}           10.43
{KW}           TRANSMEMBRANE 1
{KW}           LOW_COMPLEXITY 29.84 %

SEQ  MAAAAVTGQRPETAAAEASRPQWAPPDHCQAQAAAGLGDGEDAPVRPLCKPRGICSRAY
SEG  .....
PRD  cccceeeccccchhhhhhhhhccccccchhhhhhhccccccccccccccccccccchhhh
MEM  .....

SEQ  FLVLMVFVHLYLGNVLALLLVHYSNGDESSDPGPQHRAQGPPEPTLGPLTRLEGIKVR
SEG  .....xxxxxxxxxxxxxxxx.....
PRD  hhhhhhhhhhhhhhhhhhhhhccccccccccccccccccccccccccccccccceeeee
MEM  ....MMMMMMMMMMMMMMMM.....

SEQ  TSLPRRAPGPAPRLSPRGPALSPGPHAAPGAALRRSRALPLTRLLSLSGPGRLGPFWAARS
SEG  .....xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx
PRD  eccccccccccccccccccccccccccccccccchhhhhhhccccceccccccccchhhhhc
MEM  .....

SEQ  GAPAAARCAPPF
SEG  xxxxxxxxxx..
PRD  cccccccccc
MEM  .....
```

(No Prosite data available for DKFZphut1\_21d15.2)

(No Pfam data available for DKFZphut1\_21d15.2)

DKFZphutel\_22d2

group: signal transduction

DKFZphutel 22d2 encodes a novel 580 amino acid putative GTP-binding protein related to the ras protein. Additionally, the putative protein contains an EF-hand for calcium-binding.

G-proteins are involved in various signal transduction pathways, transferring the signal of a cellular receptor to an intracellular signal cascade.

The new protein can find clinical application in modulating/blocking the response to a cellular receptor.

similarity to GTP-binding proteins

complete cDNA, complete cds, potential start at Bp 64, EST hits  
complete cds according to K08F11.5 and YAL048c

Sequenced by BMFZ

Locus: /map="17"

Insert length: 3247 bp

Poly A stretch at pos. 3230, no polyadenylation signal found

```

1 CTCTCTGGTGA GAGGAGTCCA CTCCGTGCGT GCGGGCGGAG GCCGGCCCCC
51 GAGAGCCGCC GACATGAAGA AAGACGTGCG GATCCTGCTG GTGGGAGAAC
101 CTAGAGTTGG GAAGACATCA CTGATTATGT CTCTGGTCAG TGAAGAATTT
151 CCAGAAAGAGG TTCTCTCCCG GGCAGAAAGAA ATCACCATT CAGCTGATGT
201 CACCCAGAG AGAGTTCCAA CACACATTGT AGATTACTCA GAAGCAGAAC
251 AGAGTGATGA ACAACTTCAT CAAGAAATAT CTCAGGCTAA TGTCTCTGT
301 ATAGTGTATG CCGTTAACAA CAAGCATTCT ATTGATAAGG TAACAAGTCG
351 ATGGATTCCCT CTCATAAATG AAAGAACAGA CAAAGACAGC AGGCTGCCTT
401 TAATATTGGT TGGGAACAAA TCTGATCTGG TGGAATATAG TAGTATGGAG
451 ACCATCCTTC CTATTATGAA CCAGTATACA GAAATAGAAA CCTGTGTGGA
501 GTGTTGAGCG AAAAACCTGA AGAACATATC AGAGCTCTTT TATTACGCAC
551 AGAAAGCTGT TCTTCATCCT ACAGGGCCCC TGACTGCCC AGAGGAGAAG
601 GAGATGAAGC CAGCTTGTAT AAAAGCCCTT ACTCGTATAT TTAATAATATC
651 TGATCAAGAT AATGATGGTA CTCTCAATGA TGCTGAACCT AACTTCTTTC
701 AGAGGATTTG TTTCAACACT CCATTAGCTC CTCAGCTCTT GGAGGATGTC
751 AAGAATGTAG TCAGAAAACA TATAAGTGAT GGTGTGGCTG ACAGTGGGTT
801 GACCCGTGAA GGTTTTCTCT TTTTACACAC ACTTTTATAT CAGAGAGGGA
851 GACACGAAC TACTTGGACT GTGCTTCGAC GATTGGTTA TGATGATGAC
901 CTGGATTGTA CACCTGAATA TTTGTTCCCC CTGCTGAAAA TACCTCTGTA
951 TTGCACTACT GAATTAATC ATCATGCATA TTTATTTCTC CAAAGCACCT
1001 TTGCAAGCA TGATTGGAT AGAGACTGTG CTTTGTCACT TGATGAGCTT
1051 AAGATTATG TTAAGTTTT CCCTTACATA CCTTGGGGGC CAGATGTGAA
1101 TAACACAGTT TGTACCAATG AAAGAGGCTG GATAACCTAC CAGGGATTCC
1151 TTTCCAGTG GACGCTCAGC ACTTATTTAG ATGTACAGCG GTGCTTGGAA
1201 TATTTGGGCT ATCTAGGCTA TTCAATATTG ACTGAGCAAG AGTCTCAAGC
1251 TTGAGCTGTT ACAGTGACAA GAGATAAAAA GATAGACCTG CAGAAAAAAC
1301 AAATCTCAAG AAATGTGTTT AGATGTAATG TAATTGGAGT GAAAACTGT
1351 GGGAAAAGTG GAGTCTTCTA GGCTCTTCTT GGAAGAACT TAATGAGGCA
1401 GAAGAAAATT CGTGAAGATC ATAAATCCTA CTATGCGATT AACACTGTTT
1451 ATGTATATGG ACAAGAGAAA TACTTGTGTT TGCATGATAT CTCAGAATCG
1501 GAATTTCTAA CTGAAGCTGA AATCATTTGT GATGTTGTAT GCCTGGTATA
1551 TGATGTCAGC AATCCCAAT CCTTTGAATA CTGTGCCAGG ATTTTAAAGC
1601 AACACTTTAT GGACAGCAGA ATACCTTGCT TAATCGTAGC TGCAAAGTCA
1651 GACCTGCATG AAGTTAAACA AGAATACAGT ATTTACCTTA CTGATTCTGT
1701 CAGGAACAC AAAATGCCTC CACCACAAGC CTTCACTTGC AATACTGCTG
1751 ATGCCCCCAG TAAGGATATC TTTGTTAAAT TGACAACAAT GGCCATGTAT
1801 CCGTAAGTAC TTGCTGTCTT CATTTTCATG TTGATGGTT CATAACATTG
1851 CATGCCATTA TTAGCCATGA AGGGAATATC TTTGTCACAT AGGAATTGTT
1901 CAGCAACAGA AAGATACTTT GTAATGAGAA GGTACAAAT TGAGTAAATG
1951 CAAGTTTGGT TTGAATGCCA TAATAAAATG ATATAAACAG TGCTTCTGAC
2001 AATATCTGTA TATTTTGTAG CAGGCTGTAA CTATCTTAAT AGAATAGTAC
2051 AATAAAACAC AACCCCCAC CCAGCATTA AAAATAGTTT TACTGGAATA
2101 AAATGGGTTT GGCATCATGT TGTTTTATGC TTATAAGCA TTTTCATATG
2151 AACAGAAAGT TTATATTTT CTGTTTTTGA CCTTAGGTAT ATGAAGTTTT
2201 CTAATAATTT TTATTAATTT ATGTTGAAAT TGTGGGTATG CTTCACTTAG
2251 GATATGTCTT TTTAAGTGC TGTAAGAGT AGTTGTAATT GGAATTTCTA
2301 CTGTATAAAT GTTTTACATT AAGTGTTACG AGCCACAAT TTCAATGTACA
2351 TTTATTATAT ATCTATACAT GCATATGCAC AAGCACATAA CTGTGGTCAT
2401 CTCTGTAGTT TACTAACTGC CTTAAATTTG CATGTTCTT AATGGCATTC
2451 GCCTCAAGTA GTGTGTTTGT ATAAATCTG TTTTGTAAAC AAATAGTTTT
2501 TCAGGCAAGT CGTTTCTCAG GACTTTATAG CTTATTCTAC TTATTCTTAT
2551 GTTAGTCTCT AAATTATTTT TCTTCTTATG AAAACTACAG TGTAACACAG

```

```

2601 AGTAATAATC AACATTGCT ATAAACCAAG AATGACATTT TTCAAAAAGG
2651 TGTGTGATTG TACAGATTTT TAAAGTCAGT TAACTTTACT GCTATTTTAT
2701 TACCTAATAC TTTTTTTAGA TGCAACAAAC CCTTGAATTT CTATTTGTAT
2751 TCGAAGACAA GTCATTCTTA TTATTATAGA ATAACCAAAA CCTTATTAT
2801 GTTTTACCTT TGCTTTAAAA CTCTCATGTA TGTATCTAC AGAGAGGATC
2851 ATTACAGAGA CAGACTCTCC CGAGACATGG GCCACACTGA TAGAATAGAG
2901 AATTTGAGAA AAATCTGGGT CTTTCTAAAA ACTGCTTTGT AAGTTACTTT
2951 TTCTTTATGA CTCTGTGGG ATTTTGTGA TATTTCTTA GAGAATGACC
3001 AAATCTCCTT TCTTGCCATA ATTAACATTT AGTAATTATG TAGAAACGCA
3051 CTGCTTGGTC AGGCTTCCTG CCTAGCTATA TATTACGTTG TCTTCCTTAC
3101 TACATAAATG TACTTCTTAA ATCTTGTGAT TACAGTAAC GCAAGTGTGT
3151 TTTTACATCT GCATTTTAA AACATTTTAC TGTAATTCTG TTGTGTGTGT
3201 GTGTGTATA TGATAAATGT ACATACATGG AAAAAAAAAA AAAAAA

```

## BLAST Results

-----

Entry AC004527 from database EMBL:  
 \*\*\* SEQUENCING IN PROGRESS \*\*\* NF1-related locus, Direct Submission;  
 HTGS phase 1, 10 unordered pieces.  
 Score = 1899, P = 1.1e-78, identities = 387/396

Entry HS148355 from database EMBL:  
 human STS SHGC-31220.  
 Score = 1826, P = 7.5e-78, identities = 388/406

## Medline entries

-----

No Medline entry

## Peptide information for frame 1

-----

ORF from 64 bp to 1803 bp; peptide length: 580  
 Category: similarity to known protein

```

1 MKKDVRILLV GEPRVGKTSI IMSLVSEEF EEVPPRAEEI TIPADVTPE
51 VPTIIVDYSE AEQSDQLHQ EISQANVICI VYAVNNKHSI DKVTSRWIPL
101 INERTDKDSR LPLILVGNKS DLVEYSSMET ILPINNQYTE IETCVCESAK
151 NLKNISELFY YAKAVLHPT GLYCPPEKE MKPACIKALT RIFKISDQDN
201 DGTILNDAELN FFQRICFNTP LAPQALDVK NVVRKHISDG VADSGTLKG
251 FLFLHTLFIQ RGRHETWTV LRRFGYDDL DLTPEYLFPL LKIPPDCTTE
301 LNHAYLFLQ STFDKHLDR DCALSPDELK DLFKVPYIP WGPDVNNTVC
351 TNERGWITYQ GFSLQWTLTT YLDVQRCLEY LGYLGYSILT EQESQASAVT
401 VTRDKKIDIQ KKQTRNVFR CNVIGVKNCG KSGVLQALLG RNLMRQKKIR
451 EDHKSYYAIN TVYVYQEKY LLLHDISESE FLTEAEIICD VVCLVYDVSN
501 PKSFEYCARI FKQHFMSRI PCLIVAAKSD LHEVKQEYSI SPTDFCRKHK
551 MPPQAFTCN TADAPSKDIF VKLTTMAMYP

```

## BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphut1\_22d2, frame 1

TREMBL:CEUK08F11\_3 gene: "K08F11.5"; Caenorhabditis elegans cosmid  
 K08F11., N = 1, Score = 1357, P = 1.1e-138

TREMBL:SPCC320\_4 gene: "SPCC320.04c"; product: "hypothetical protein";  
 S.pombe chromosome III cosmid c320., N = 1, Score = 889, P = 4.4e-89

TREMBL:CEUC47C12\_3 gene: "C47C12.4"; Caenorhabditis elegans cosmid  
 C47C12., N = 2, Score = 408, P = 5.6e-74

PIR:S51971 probable membrane protein YAL048c - yeast (Saccharomyces  
 cerevisiae), N = 1, Score = 677, P = 1.3e-66

>TREMBL:CEUK08F11\_3 gene: "K08F11.5"; Caenorhabditis elegans cosmid  
 K08F11.  
 Length = 625

HSPs:



Score = 1357 (203.6 bits), Expect = 1.1e-138, P = 1.1e-138  
Identities = 263/582 (45%), Positives = 380/582 (65%)

```

Query:      4 DVRILLVGEPRVGKTSLSIMSLVSEEFPEEVPPRAEEITIPADVTPERVPTTHIVDYSEAEQ 63
             DVRI+L+G+  GKTS+MSL+ +E+ + VP R + + IPADVTP E V T IVD S  E+
Sbjct:      9 DVRIVLIGDEGCGKTSLSVMSLLEDEWVDVAVPRRLDRVLIPADVTPENVTTTSIVDLISKEE 68

Query:     64 SDEQLHQEISQANVICIVYAVNNKHSIDKVTSRWIPLINERTDKDSRLPLILVGNKSDLV 123
             + + EI QANVIC+VY+V ++ +D + ++W+PLI + + P+ILVGNKSD
Sbjct:     69 DENWIVSEIRQANVICVYVSVTDESTVDGIQTKWLPILIRQSFGEYHETPVILVGNKSDGT 128

Query:    124 EYSSMETILPIMNOYTEIETCVECSAKNLKNISELFYYAQKAVLHPTGPLYCPEEKEMKP 183
             ++ + ILPIM TE+ETCVECSA+ +KN+SE+FYAQKAV++PT PLY + K++
Sbjct:    129 A-NNTDKILPIMEANTEVETCVECSARTMKNVSEIFYAQKAVIYPTRPPLYDADTKQLTD 187

Query:    184 ACIKALTRIFKISDQNDGTLNDAELNFFQRICFNTPLAPQAEDEVKNVVRKHISDGVAD 243
             KAL R+FKI D+DNDG L+D ELN FQ++CF PL ALEDVK V DGVA+
Sbjct:    188 RARKALIRVFKICDRDNDGYLSOTELNDFQKLCFGIPLTSTALEDVKRAVSDGCPDGVAN 247

Query:    244 SGLTLKGFLFLHTLFIQRGRHETTTWTLRRFGYDDDLDTPEYLFPLLKIPPDCTTELNH 303
             L L GFL+LH LFI+RGRHETTW VLR+FGY+ L L+ +YL+P + IP C+TEL+
Sbjct:    248 DSLMLAGFLYLHLLFIERGRHETTWAVLRKFGYETSLKLSYDLYPRITIPVGCSTELSP 307

Query:    304 HAYLFLQSTFDKHDLDRCALSPDELKDLFKVFPYIPWGPVNNVTCTNERGWITYQGFL 363
             F+ + F+K+D D+D LSP EL++LF V P D + TN+RGW+TY G++
Sbjct:    308 EGVQFVSALFEKYDEDKDGLSPSELQNLFSVCPVPVITKDNILAETNQRGWLTYNNGYM 367

Query:    364 SQWTLTLYLDVQRCLEYLGYLGSILTEQESQAS----AVTTRDKKIDLQKKQTQRNVF 419
             + W +TT +++ + E L YLG+ + +A ++ VTR++K DL+ T R VF
Sbjct:    368 AYNNMTTLINLTQTFEQLAYLGFPPVGRSGPGRAGNTLDSIRVTRERKKDLENHGTDRKVF 427

Query:    420 RCNVIGVKNCGSGVLQALLGRNLMRQKKIREDHKSYYAINTVYVYGQEKYLLLDHI--- 476
             +C V+G K+ GK+ +Q+L GR + +I H S + IN V V + KYLL ++
Sbjct:    428 QCLVVGAAGDAGKTVFMQSLAGRGMAADVAGIGRRH-SPFVINRVRVKESKYLLLRVDVL 486

Query:    477 SESEFLTEAEIICDVVCLVYDVSNPKSFEYCARIFKQHFMDSRIPCLIVAASDLHEVKQ 536
             S + L E DVV +YD+SNP SF +CA +++++F ++ PC+++A K + EV Q
Sbjct:    487 SPQDALGSGETSADVVAFLYDISNPDSFAFCATVYQYFYRTKTPCVMIATKVEREEVDQ 546

Query:    537 EYSISPTDFCRKHKMPPPPQAFTCNTADAPSKDIFVKLTMMAMP 580
             + + P +FCR+ ++P P F+ S IF +L MA+YP
Sbjct:    547 RWEVPPEEFCRQFELPKPIKFSTGNIGQSSSPIFEQLAMMAVYP 590

```

Pedant information for DKFZphutel\_22d2, frame 1

#### Report for DKFZphutel\_22d2.1

```

[LENGTH]      580
[MW]           66541.61
[pI]           5.56
[HOMOL]        TREMBL:CEUK08F11_3 gene: "K08F11.5"; Caenorhabditis elegans cosmid K08F11. 1e-
149
[FUNCAT]       99 unclassified proteins [S. cerevisiae, YAL048c] 5e-81
[FUNCAT]       03.04 budding, cell polarity and filament formation [S. cerevisiae, YKR055w]
3e-11
[FUNCAT]       03.99 other cell growth, cell division and dna synthesis activities [S.
cerevisiae, YNL098c] 8e-09
[FUNCAT]       10.04.07 g-proteins [S. cerevisiae, YNL098c] 8e-09
[FUNCAT]       03.10 sporulation and germination [S. cerevisiae, YNL098c] 8e-09
[FUNCAT]       11.01 stress response [S. cerevisiae, YNL098c] 8e-09
[FUNCAT]       03.22 cell cycle control and mitosis [S. cerevisiae, YNL098c] 8e-09
[FUNCAT]       01.03.13 regulation of nucleotide metabolism [S. cerevisiae, YNL098c]
8e-09
[FUNCAT]       01.05.04 regulation of carbohydrate utilization [S. cerevisiae, YNL098c]
8e-09
[FUNCAT]       30.03 organization of cytoplasm [S. cerevisiae, YOR101w] 4e-08
[FUNCAT]       11.10 cell death [S. cerevisiae, YOR101w] 4e-08
[FUNCAT]       10.02.07 g-proteins [S. cerevisiae, YPR165w] 7e-08
[FUNCAT]       30.04 organization of cytoskeleton [S. cerevisiae, YPR165w] 7e-08
[FUNCAT]       30.08 organization of golgi [S. cerevisiae, YPR165w] 7e-08
[FUNCAT]       08.07 vesicular transport (golgi network, etc.) [S. cerevisiae, YFL005w]
9e-08
[FUNCAT]       30.09 organization of intracellular transport vesicles [S. cerevisiae,
YFL005w] 9e-08
[FUNCAT]       30.02 organization of plasma membrane [S. cerevisiae, YFL005w] 9e-08
[FUNCAT]       08.13 vacuolar transport [S. cerevisiae, YNL093w] 1e-07

```

```

[FUNCAT]      06.04 protein targeting, sorting and translocation [S. cerevisiae, YNL093w]
1e-07
[FUNCAT]      08.19 cellular import [S. cerevisiae, YNL093w] 1e-07
[FUNCAT]      10.05.07 g-proteins [S. cerevisiae, YLR229c] 8e-07
[FUNCAT]      03.07 pheromone response, mating-type determination, sex-specific proteins
[S. cerevisiae, YLR229c] 8e-07
[FUNCAT]      10.99 other signal-transduction activities [S. cerevisiae, YCR027c] 3e-06
[FUNCAT]      09.09 biogenesis of intracellular transport vesicles [S. cerevisiae,
YGL210w] 9e-04
[BLOCKS]      BL00410A Dynamin family proteins
[SCOP]         dlp1k_ 3.25.1.3.1 cH-p21 Ras protein [human (Homo sapiens)] 2e-42
[SCOP]         d1guaa_ 3.25.1.3.10 RapiA [Human (Homo sapiens)] 5e-59
[PIRKW]        transmembrane protein 1e-79
[PIRKW]        membrane trafficking 2e-06
[PIRKW]        acetylated amino end 3e-09
[PIRKW]        prenylated cysteine 3e-09
[PIRKW]        signal transduction 1e-07
[PIRKW]        transforming protein 3e-09
[PIRKW]        immediate-early protein 8e-06
[PIRKW]        alternative splicing 4e-08
[PIRKW]        P-loop 1e-10
[PIRKW]        lipoprotein 7e-10
[PIRKW]        proto-oncogene 3e-09
[PIRKW]        methylated carboxyl end 3e-09
[PIRKW]        membrane protein 3e-09
[PIRKW]        GTP binding 1e-10
[PIRKW]        thiolester bond 7e-10
[SUPFAM]       ras transforming protein 1e-10
[PROSITE]      ATP_GTP_A 2
[PROSITE]      MYRISTYL 3
[PROSITE]      EF_HAND 1
[PROSITE]      CAMP_PHOSPHO_SITE 1
[PROSITE]      CK2_PHOSPHO_SITE 14
[PROSITE]      TYR_PHOSPHO_SITE 4
[PROSITE]      PKC_PHOSPHO_SITE 5
[PROSITE]      ASN_GLYCOSYLATION 3
[PFAM]         Ras family (contains ATP/GTP binding P-loop)
[KW]           Irregular
[KW]           3D

```

```

SEQ      MKKDVRILLVGEPRVGKTSLIMSLVSEEFPEEVPPRAEITIPADVTPERVPTHIVDYSE
1jai-    ...EEEEEEEEETTTCHHHHHHHHHHCCCCCCCCCEEEEEETEEEEEEEEEECCC

```

```

SEQ      AEQSDQLHQEISQANVICIVAVNNKHSIDKVTSRWIPLINERTOKDSRLPLILVGNKS
1jai-    CGGGHHHHHHHHHHTTEEEEEETTHHHHHHHH-HHHHHHHHHHCTTT-TCEEEEEETT

```

```

SEQ      DLVEYSSMETILPIMNQYTEIETCVCESAKNLKNISELFYYAQAVLHPTGPLYCPREEKE
1jai-    TTTTTTTHHHHHHHHHHHHCCCE-EECTTTTTTHHHHHH.....

```

```

SEQ      MKPACIKALTRIFKISDQDNDGTLNDAELNFFQRICFNTPLAPQALVDKNNVVRKHISDG
1jai-    .....

```

```

SEQ      VADSGTLTKGFLFLHTLFIQGRHETTWTVLRRFGYDDDLDTPEYLFPLKIPDCTTE
1jai-    .....

```

```

SEQ      LNHAYLFLQSTFDKHDLDRCALSPDELKDLFKVFPYIPWGPVNNVTCTNERGWITYQ
1jai-    .....

```

```

SEQ      GFLSQWTLTTYLDVQRCLEYLGYLGYSILTEQESQASAVTVTRDKKIDLQKKQTQRNVFR
1jai-    .....

```

```

SEQ      CNVIGVKNCGSGVLQALLGRNLMRQKKIREDHKSYYAINTVYVYGQEKYLLHDISESE
1jai-    .....

```

```

SEQ      FLTEAEIICDVVLVDVSNPKSFEYCARIFKQHFMDSRIPCLIVAAKSDLHEVKQEYSI
1jai-    .....

```

```

SEQ      SPTDFCRKHKMPPQAFTCNTADAPSKDIFVKLTMMAMP
1jai-    .....

```

#### Prosite for DKFZphut1\_22d2.1

PS00001	118->122	ASN_GLYCOSYLATION	PDOC00001
PS00001	154->158	ASN_GLYCOSYLATION	PDOC00001
PS00001	346->350	ASN_GLYCOSYLATION	PDOC00001
PS00004	411->415	CAMP_PHOSPHO_SITE	PDOC00004
PS00005	94->97	PKC_PHOSPHO_SITE	PDOC00005
PS00005	105->108	PKC_PHOSPHO_SITE	PDOC00005

PS00005	148->151	PKC_PHOSPHO_SITE	PDOC00005
PS00005	247->250	PKC_PHOSPHO_SITE	PDOC00005
PS00005	414->417	PKC_PHOSPHO_SITE	PDOC00005
PS00006	59->63	CK2_PHOSPHO_SITE	PDOC00006
PS00006	105->109	CK2_PHOSPHO_SITE	PDOC00006
PS00006	126->130	CK2_PHOSPHO_SITE	PDOC00006
PS00006	139->143	CK2_PHOSPHO_SITE	PDOC00006
PS00006	143->147	CK2_PHOSPHO_SITE	PDOC00006
PS00006	196->200	CK2_PHOSPHO_SITE	PDOC00006
PS00006	203->207	CK2_PHOSPHO_SITE	PDOC00006
PS00006	311->315	CK2_PHOSPHO_SITE	PDOC00006
PS00006	325->329	CK2_PHOSPHO_SITE	PDOC00006
PS00006	370->374	CK2_PHOSPHO_SITE	PDOC00006
PS00006	390->394	CK2_PHOSPHO_SITE	PDOC00006
PS00006	477->481	CK2_PHOSPHO_SITE	PDOC00006
PS00006	483->487	CK2_PHOSPHO_SITE	PDOC00006
PS00006	541->545	CK2_PHOSPHO_SITE	PDOC00006
PS00007	153->161	TYR_PHOSPHO_SITE	PDOC00007
PS00007	376->384	TYR_PHOSPHO_SITE	PDOC00007
PS00007	153->162	TYR_PHOSPHO_SITE	PDOC00007
PS00007	448->457	TYR_PHOSPHO_SITE	PDOC00007
PS00008	240->246	MYRISTYL	PDOC00008
PS00008	425->431	MYRISTYL	PDOC00008
PS00008	433->439	MYRISTYL	PDOC00008
PS00017	11->19	ATP_GTP_A	PDOC00017
PS00017	425->433	ATP_GTP_A	PDOC00017
PS00018	197->210	EF_HAND	PDOC00018

## Pfam for DKFZphut1\_22d2.1

HMM_NAME	Ras family (contains ATP/GTP binding P-loop)		
HMM	*KLVLLIGDSGVGKSCLLIRFTQNeFnEeYIPTIGvDFYtKTIEIDGktIK		
	++L+G+ VGK++L ++ EF+EE +P ++ T ++ +++		
Query	6	RILLVGEPRVGKTSLSIMSLVSEEFPEE-VPPR-AEEITIPADVTPERVP	52
HMM	LQIWDTAGQERYRsmRPMYYRGAMGFMVLYDITNRqSFENIr.NWweEIr		
	I D E+ + + +A+++ +VY+++N+ S ++++ +W++ I+		
Query	53	THIVDYSEAEQSDQLHQEISQANVICIVYAVNNKHSIDKVTSRWIPLIN	102
HMM	RHCDrDENVPIMLVGNKCDLEDQRQVStEEGQeFAREWGAIPFMETSAKT		
	+ D+D+ P +LVGNK+DL + ++T + +E+SAK+		
Query	103	ERTDKDSRLPLILVGNKSDLVEYSSMETILPIMNQYTEI-ETCVECSAKN	151
HMM	NINVEEAFMEIvReIlqrMqeqNqteNinidQpsrnrkrCCCIM*		
	N+ E F+ + +++L + . +++ +++++ + C+		
Query	152	LKNISELFYYAQKAVLHPT-----GLYCPEEKEMK-PACI--	186

DKFZphute1\_22e12

group: signal transduction

DKFZphute1\_22e12 encodes a novel 92 amino acid protein, with similarity to yeast, C.elegans, Drosophila and mammalian proteins.

The Drosophila cni and mammalian cornichon proteins are part of a signal transduction pathway involving the EGF-receptor.

The new protein can find application in modulating the cornichon modulated signal transduction way and also the EGF receptor signaling processes.

strong similarity to S.cerevisiae YGL054c and cornichon

complete cDNA, complete cds, EST hits  
cornichon is required for signal transduction in the EGF-receptor  
signal processing

Sequenced by BMFZ

Locus: unknown

Insert length: 519 bp

Poly A stretch at pos. 499, no polyadenylation signal found

```

1  GTCGGGGCAT CCGAGCGGGT TTGACGGAAG GAGCGGCGGC GACGGAGGAG
51 GAGGATGGAG GCGGTGGTGT TCGTCTTCTC TCTCCTCGAT TGTTCGCGGC
101 TCATCTTCCT CTCGGTCTAC TTCATAATTA CATTGTCTGA TTTAGAATGT
151 GATTACATTA ATGCTAGATC ATGTTGCTCA AAATTAAACA AGTGGGTAAT
201 TCCAGAATTG ATTGGCCATA CCATTGTCAC TGTATTACTG CTCATGTCAT
251 TGCACCTGGT CATCTTCCTT CTCAACTTAC CTGTTGCCAC TTGGAATATA
301 TATCGTATGA TCTTAGCTTT GATAAATGAC TGAAGCTGGA GAAGCCGTGG
351 TTGAAGTCAG CCTACACTAC AGTGCACAGT TGAGGAGCCA GAGACTTCTT
401 AAATCATCCT TAGAACCGTG ACCATAGCAG TATATATTTT CCTCTTGGA
451 CAAAAAACTA TTTTGCTGT ATTTTACCA TATAAGTAT TAAAAAACA
501 TGAAAAAATA AAAAAAATA

```

## BLAST Results

No BLAST result

## Medline entries

95300228:  
cornichon and the EGF receptor signaling process are necessary for both  
anterior-posterior  
and dorsal-ventral pattern formation in Drosophila.

## Peptide information for frame 1

ORF from 55 bp to 330 bp; peptide length: 92  
Category: strong similarity to known protein

```

1 MEAVVFVFSL LDCCALIFLS VYFIITLSDL ECDYINARSC CSKLNKQVIP
51 ELIGHTIVTV LLLMSLHWFI FLLNLPVATW NIYRMILALI ND

```

## BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphute1\_22e12, frame 1

PIR:S64058 probable membrane protein YGL054c - yeast (Saccharomyces cerevisiae), N = 2, Score = 185, P = 5.7e-17

TREMBL:SPAC2C4\_5 gene: "SPAC2C4.05"; product: "cornichon homolog";

S.pombe chromosome I cosmid c2C4., N = 1, Score = 163, P = 3.7e-12

PIR:S46084 probable membrane protein YBR210w - yeast (Saccharomyces cerevisiae), N = 1, Score = 162, P = 4.8e-12

TREMBL:AF104398.1 product: "cornichon"; Homo sapiens cornichon mRNA, complete cds., N = 1, Score = 141, P = 8e-10

SWISSPROT:CNI\_DROVI CORNICHON PROTEIN., N = 1, Score = 139, P = 1.3e-09

>PIR:S64058 probable membrane protein YGL054c - yeast (Saccharomyces cerevisiae)  
Length = 138

#### HSPs:

Score = 185 (27.8 bits), Expect = 5.7e-17, Sum P(2) = 5.7e-17  
Identities = 35/85 (41%), Positives = 56/85 (65%)

Query: 1 MEAVVVFVSLDCCALIFLSVYFIITLSDLECDYINARSCSKLNKWWIPELIGHTIVTV 60  
M A +F+ +++ C +F V+F I +DLE DYIN CSK+NK + PE H +++  
Sbjct: 1 MGAWLFILAVVVCINLFGQVHFTILYADLEADYINPIELCSKVNLITPEAALHGALS 60  
Query: 61 LLLMSLHWFIFLLNLPVATWNIYRM 85  
L L++ +WF+FLLNLPV +N+ ++  
Sbjct: 61 LFLNGYWFVFLNLPVLAYNLNKI 85

Score = 37 (5.6 bits), Expect = 5.7e-17, Sum P(2) = 5.7e-17  
Identities = 7/9 (77%), Positives = 9/9 (100%)

Query: 82 IYRMILALI 90  
+YRMI+ALI  
Sbjct: 123 LYRMIMALI 131

#### Pedant information for DKFZphutel\_22e12, frame 1

#### Report for DKFZphutel\_22e12.1

[LENGTH] 92  
[MW] 10614.98  
[pI] 5.04  
[HOMOL] PIR:S64058 probable membrane protein YGL054c - yeast (Saccharomyces cerevisiae)  
5e-14  
[FUNCAT] 03.04 budding, cell polarity and filament formation [S. cerevisiae, YGL054c]  
2e-15  
[PIRKW] transmembrane protein 2e-11  
[PROSITE] CK2\_PHOSPHO\_SITE 3  
[KW] SIGNAL PEPTIDE 33  
[KW] TRANSMEMBRANE 2

SEQ MEAVVVFVSLDCCALIFLSVYFIITLSDLECDYINARSCSKLNKWWIPELIGHTIVTV  
PRD ccchhhhhhhhhhhhhhhhhheeeccccccccccccccccccccceehhhhhhhhhhhhh  
MEM .....MMMMMMMMMM

SEQ LLLMSLHWFIFLLNLPVATWNIYRMILALIND  
PRD hhhhhhhheeeccccchhhhhhhhhhhhhcc  
MEM MMMMMMMMMMMMMMMMM..MMMMMM....

#### Prosite for DKFZphutel\_22e12.1

PS00006 9->13 CK2\_PHOSPHO\_SITE PDOC00006  
PS00006 26->30 CK2\_PHOSPHO\_SITE PDOC00006  
PS00006 28->32 CK2\_PHOSPHO\_SITE PDOC00006

(No Pfam data available for DKFZphutel\_22e12.1)

DKFZphut1\_22n2

group: uterus derived

DKFZphut1\_22n2 encodes a novel 304 amino acid protein without similarity to known proteins.

No informative BLAST results; No predictive prosite, pfam or SCOP motif.

The new protein can find application in studying the expression profile of uterus-specific genes.

unknown

complete cDNA, complete cds, EST hits

Sequenced by BMFZ

Locus: /map="553.3 cR from top of Chr11 linkage group"

Insert length: 1556 bp

Poly A stretch at pos. 1534, no polyadenylation signal found

```
1 ACAACAGGCT GGTGCTTGG CGTGGAATCC TAAAGTGCC TGGCTTTGAG
51 ACTGGAGTGA GACCCAGCC CTAGGCTGGG GTTCTTTCCA TTATAGAGGA
101 GACGGATTCA GAAGGGCTAC AGACCAAGGT TGTGAAAAC CAGACATATG
151 ATGAGCGTCT AGAGATTAA GACTCCGAAG AGGTTGCAAG TATTTATCT
201 CCAACCCCAA GACACCAAGG ACTTCCTCGT TCTGCCATC TTCCTAACAA
251 GGCTATGGCT GATAACAGCA GTGATGAGTG TGAAGAGGAA AATAACAAGG
301 AGAAGAAGAA GACCTCACAG TTGACACCTC AACGGGGCTT TAGTGAAAT
351 GAGGATGACG ATGATGATGA TGATGATTCA TCTGAACTG ATTCTGATTC
401 TGATGATGAT GATGAAGAGC ATGGAGCCCC TCTGGAAGGG GCCTATGACC
451 CTGCAGACTA TGAGCATTG CCAGTTTCTG CTGAAATTAA GGAACCTCTC
501 CAGTACATCA GTAGGTACAC ACCTCAGTTG ATTGACCTGG ACCACAACT
551 GAAGCCCTTC ATTCTGATT TTATCCCAGC TGTCGGGGAT ATTGATGCAT
601 TCTTAAAGGT CCCACGTCCT GATGAAAGC CTGACAACCT TGGCCTATTG
651 GTATTGGATG AACCTTCTAC AAAGCAGTCA GACCCACGG TGCTCTCACT
701 CTGTTTAAAC GAGAATTCTA AGCAGACAA CATCACAAA CATATGAAAG
751 TAAAAAGCCT AGAAGATGCA GAAAAGAATC CCAAAGCCAT TGACACGTGG
801 ATTGAGAGCA TCTCTGAATT ACACCGTTCT AAGCCCCCTG CGACTGTGCA
851 CTACACCAGG CCCATGCCCG ACATTGACAC GCTGATGCAG GAATGGTCCC
901 CGGAGTTTGA AGAGCTTTTG GGCAAGGTAA GCCTGCCAC GGCAGAGATT
951 GATTGCAGCC TGGCAGAGTA CATTGACATG ATCTGTGCCA TTCTAGACAT
1001 CCCTGTCTAC AAGAGTCGGA TCCAGTCCCT CCATCTGCTC TTTTCCCTCT
1051 ACTCAGAATT CAAGAACTCA CAGCATTTTA AAGCTCTCGC TGAAGGCAAG
1101 AAAGCATTCA CTCCTTCATC CAATTCCACC TCCCAAGCTG GAGACATGGA
1151 GACATTAACC TTCAGCTGAG ACACCTCCCA AGCTGCTGTT TCAAGGCTGA
1201 GCTGGCCCCC CTGCCCCAGC TGAGATGGAC AGATCGTTGT CAGCTACTTG
1251 ATGTCCTTGC CCATGCCACA GCTTGGCTCA GGGGCAGTGC ATGTCCTGCT
1301 GCCCTCTCTG CCAGAGGGCA CAGAACATGT TTGTTTAATG AACCTGCCTG
1351 CCTCAGATTG CTGTCCCCGG GGAGTTAATG CATCTACACC ACTGTGGGGA
1401 TTTGAGTTAT AAGAATTGGA ATTTCTGAGA TCCCATGGAG GTTAGATTGG
1451 GAGGAAAGCT TAAAAGATGT CCTTTTGTG AGAGGGATGG AATTGTTTTT
1501 TTTCAATTCG AAAGTTAGTG AGTAAAGATT TTATAAATCA AAAAAAAAAA
1551 AAAAAA
```

## BLAST Results

Entry HS188252 from database EMBL:  
human STS WI-12265.  
Score = 2554, P = 4.1e-109, identities = 556/587

## Medline entries

No Medline entry

## Peptide information for frame 3

ORF from 255 bp to 1166 bp; peptide length: 304  
Category: putative protein

```

1  MADNSSDECE EENNKEKKKT SQLTPQRGFS ENEDDDDDDD DSSETDSDD
51 DDDDEEHGAPL EGAYDPADYE HLPVSAEIKE LFQYISRYTP QLIDLHKLK
101 PFIPDFIPAV GVIDAFLKVP RPDGKPDNLG LLVLDEPSTK QSDPTVLSLW
151 LTENSKQHNI TQHKMKVKSLE DAEKNPKAID TWIESISELH RSKPPATVHY
201 TRPMPDIDTL MQEWSPEFEE LLGKVSLPTA EIDCSLAEYI DMICAILDIP
251 VYKSRIQSLH LLFSLYSEFK NSQHFKALE GKKAFTPSSN STSQAGDMET
301 LTFS

```

## BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphutel\_22n2, frame 3

PIR:S38149 SIS2 protein - yeast (Saccharomyces cerevisiae), N = 1,  
Score = 132, P = 1e-05

>PIR:S38149 SIS2 protein - yeast (Saccharomyces cerevisiae)  
Length = 562

## HSPs:

Score = 132 (19.8 bits), Expect = 1.0e-05, P = 1.0e-05  
Identities = 24/63 (38%), Positives = 35/63 (55%)

```

Query:   3  DNSSDECEENNKEKKKTSQLTPQRGFSENEDDDDDDDSSETDSDDDDDEEHGAPLEG 62
          +  DE EEE++ E++ T          +++DDDDDDDD + D D DDD++E A  G
Sbjct:  497 EEDDDDEDEEEDDEEEDTEDKNENNNDDDDDDDDDDDDDDDDDDDEDEDEAETPG 556

Query:   63  AYD 65
          D
Sbjct:  557 IID 559

```

Score = 122 (18.3 bits), Expect = 1.4e-04, P = 1.4e-04  
Identities = 20/52 (38%), Positives = 33/52 (63%)

```

Query:   4  NSSDECEENNKEKKKTSQLTPQRGFSENEDDDDDDDSSETDSDDDDDEE 55
          N+ +E ++E+ +E + T + + N+DDDDDDDD + D D DDD++
Sbjct:  494 NNEEEDDEDEEEDDEEEDTEDKNENNNDDDDDDDDDDDDDDDDDDDDDD 545

```

Pedant information for DKFZphutel\_22n2, frame 3

## Report for DKFZphutel\_22n2.3

```

[LENGTH]      304
[MW]           34285.85
[pI]           4.37
[PROSITE]      AMIDATION      1
[PROSITE]      CAMP_PHOSPHO_SITE  2
[PROSITE]      CK2_PHOSPHO_SITE  10
[PROSITE]      PKC_PHOSPHO_SITE   1
[PROSITE]      ASN_GLYCOSYLATION  3
[KW]           All_Alpha
[KW]           LOW_COMPLEXITY    11.84 %

```

```

SEQ  MADNSSDECEENNKEKKKTSQLTPQRGFSENEDDDDDDDSSETDSDDDDDEEHGAPL
SEG  .....xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx.....
PRD  cccccccchhhhhchhhhhccccccccccccccccccccccccccccccccccccccccc

SEQ  EGAYDPADYEHLPVSAEIKELFQYISRYTPQLIDLHKLKPFIPDFIPAVGVIDAFLKVP
SEG  .....
PRD  cccccccccccchhhhhhhhhhhhhccccccccccccccccccccccccccccccccccccc

SEQ  RPDGKPDNLGLLVLDEPSTKQSDPTVLSLWLTENSKQHNIQHKMKVKSLEDAEKNPKAID
SEG  .....
PRD  cccccccccceeeccccccccccccchhhhhccccccccccccccccccccccccchhhhh

SEQ  TWIESISELHRSKPPATVHYTRPMPDIDTLMQEWSPEFEE LLGKVSLPTAEIDCSLAEYI
SEG  .....
PRD  hhhhhhhhhhhccccccccceeeccccccccchhhhhccccchhhhhccccccccccccchhhhh

SEQ  DMICAILDIPVYKSRIQSLHLLFSLYSEFKNSQHFKALEGKKAFTPSSNSTSQAGDMET
SEG  .....

```

Prosite for DKF2phutel_22n2.3			
PS00001	4->8	ASN_GLYCOSYLATION	PDOC00001
PS00001	159->163	ASN_GLYCOSYLATION	PDOC00001
PS00001	290->294	ASN_GLYCOSYLATION	PDOC00001
PS00004	17->21	CAMP_PHOSPHO_SITE	PDOC00004
PS00004	18->22	CAMP_PHOSPHO_SITE	PDOC00004
PS00005	138->141	PKC_PHOSPHO_SITE	PDOC00005
PS00006	5->9	CK2_PHOSPHO_SITE	PDOC00006
PS00006	30->34	CK2_PHOSPHO_SITE	PDOC00006
PS00006	43->47	CK2_PHOSPHO_SITE	PDOC00006
PS00006	45->49	CK2_PHOSPHO_SITE	PDOC00006
PS00006	47->51	CK2_PHOSPHO_SITE	PDOC00006
PS00006	49->53	CK2_PHOSPHO_SITE	PDOC00006
PS00006	168->172	CK2_PHOSPHO_SITE	PDOC00006
PS00006	181->185	CK2_PHOSPHO_SITE	PDOC00006
PS00006	185->189	CK2_PHOSPHO_SITE	PDOC00006
PS00006	235->239	CK2_PHOSPHO_SITE	PDOC00006
PS00009	280->284	AMIDATION	PDOC00009

520



DKFZphut1\_22o2

group: uterus derived

DKFZphut1\_22o2 encodes a novel 537 amino acid protein without similarity to known proteins.

No informative BLAST results; No predictive prosite, pfam or SCOP motife.

The new protein can find application in studying the expression profile of uterus-specific genes.

similarity to S.pombe SPBC3E7.03c

complete cDNA, complete cds, EST hits

Sequenced by BMFZ

Locus: map="11p15.5"

Insert length: 2714 bp

Poly A stretch at pos. 2695, polyadenylation signal at pos. 2677

```
1 GCAGGGCAGC GTGGGGGCTG AGATCGTTTC CTGTTGGAAC TTCTGGCCCA
51 AGAAGCCGCG GTCAACAAGG GAGGGGTCAG TTCGGTTCAG AGCGACTCAG
101 CCCCTCGACT CGGGTCTTAA AACCTCCGAG CCGCCAGTTC TGCCTCAGGC
151 CGCGCCCCCT TAAAGCGCCA CCAGACGCTG CGCCCCGTTA AAGCGCCACC
201 AGACGCCGCG CCCCGTCCCG GCCTCCCCCG CGCGCTGGCG CGGGGCTTTC
251 TGGGCCAGGG CGGGGCCGCG GAACTGCGGC CCGGAACGGC TGAGGAAGGG
301 CCCGTCCCGC CTTCCTCCGC GCGCCATGGA GCCCGGGCG GTTGCAAGAAG
351 CCGTGAGAGC GGGTGAGGAG GATGTGATTA TGAAGCTCT GCGGTCATAC
401 AACCAAGGAG ACTCCAGAG CTTCACGTTT GATGATGCCC AACAGGAGGA
451 CCGGAAGAGA CTGGCGGAGC TGCTGGTCTC CGTCTGGAA CAGGGCTTGC
501 CACCTCCCA CCGTGTATC TGGGTGAGA GTGTCCGAAT CCTGTCCCGG
551 GACCCGAAC CTCTGGACCC GTTACCAGC CGCCAGAGCC TGCAGGCACT
601 AGCCTGCTAT GCTGACATCT CTGTCTCTGA GGGGTCCGTC CCAGAGTCCG
651 CAGACATGGA TGTGTACTG GAGTCCCTCA AGTGCTGTG CAACCTCGTG
701 CTCAGCAGCC CTGTGGCACA GATGCTGGCA GCAGAGGCC GCCTAGTGGT
751 GAAGCTCACA GAGCGTGTGG GGCTGTACCG TGAGAGGAGC TTCCCCCAGC
801 ATGTCCAGTT CTTTGACTTG CGGCTCCTCT TCCTGCTAAC GGCACCTCCG
851 ACCGATGTGC GCCAGCAGCT GTTTCAGGAG CTGAAAGGAG TGCCTGTGCT
901 AACTGACACA CTGGAGCTGA CGCTGGGGGT GACTCCTGAA GGAACCCCC
951 CACCCACGCT CCTTCCTTCC CAAGAGACTG AGCGGGCCAT GGAGATCCTC
1001 AAAGTGCTCT TCAACATCAC CCTGGACTCC ATCAAGGGGG AGGTGGACGA
1051 GGAAGACGCT GCCCTTTACC GACACCTGGG GACCTTCTCT CGGCACTGTG
1101 TGATGATGCG TACTGTGGA GACCGCAGC AGGAGTTCCA CGGCCACGCA
1151 GTGAACCTCC TGGGAACTT GCCCTCAAG TGTCTGGATG TTCTCTCAC
1201 CCTGGAGCCA CATGGAGACT CCACGGAGTT CATGGGAGTG AATATGGATG
1251 TGATTCGTGC CCTCCTCATC TTCTAGAGA AGCGTTTGCA CAAGACACAC
1301 AGGCTGAAGG AGAGTGTAGC TCCCGTGCTG AGCGTGCTGA CTGAATGTGC
1351 CCGGATGCAC CGCCAGCCA GGAAGTTCTT GAAGGCCAGG GGATGGCCAC
1401 CTCGCCAGGT GCTGCCCTCT CTGCGGGATG TGAGGACACG GCCTGAGGTT
1451 GGGGAGATGC TGCAGAACAA GCTTGCTCCG CTCATGACAC ACCTGGACAC
1501 AGATGTGAAG AGGGTGGCTG CCGAGTTCTT GTTGTCTCTG TGCTCTGAGA
1551 GTGTGCCCGG ATTCATCAAG TACACAGGCT ATGGGAATGC TGCTGGCCTT
1601 CTGGCTGCCA GGGGCCCTCAT GGCAGGAGGC CGGCCGAGG GCCAGTACTC
1651 AGAGGATGAG GACACAGACA CAGATGAGTA CAAGGAAGCC AAAGCCAGCA
1701 TAAACCTGTG GACCGGGAGG GTGGAGGAGA AGCCGCCTAA CCCTATGGAG
1751 GGCATGACAG AGGAGCAGAA GGAGCAGGAG GCCATGAAGC TGGTGACCAT
1801 GTTTGACAAG CTCTCCAGGA ACAGAGTCAT CCAGCCAATG GGGATGAGTC
1851 CCGGGGTCAT TCTTACGTCC CTGCAGGATG CCATGTGCGA GACTATGGAG
1901 CAGCAGCTCT CCTCGGACCC TGACTCGGAC CCTGACTGAG GATGGCAGCT
1951 CTCTGTCTCC CCCATCAGGA CTGGTGTGTC TTCCAGAGAC TTCTTGGGG
2001 TTGCAACCTG GGAAGCCAC ATCCCACTGG ATCCACACCC GCCCCCACTT
2051 CTCCATCTTA GAAACCCCTT CTCTTGACTC CCGTTCTGTT CATGATTGTC
2101 CTCTGGTCCA GTTCTCATC TCTGGACTGC AACGGTCTTC TTGTGCTAGA
2151 ACTCAGGCTC AGCCTCGAAT TCCACAGACG AAGTACTTTC TTTGTCTGTC
2201 GCCAAGAGGA ATGTGTTCAG AAGCTGTGTC CTGAGGGCAG GGCCTACCTG
2251 GGCACACAGA AGAGCATATG GGAGGCGAGG GGTTCGGGTG TGGGTGCACA
2301 CAAAGCAAGC ACCATCTGGG ATTGGCACAC TGGCAGAGCC AGTGTGTTGG
2351 GGTATGTGCT GCACCTCCCA GGGAGAAAC CTGTGAGAAC TTTCATACG
2401 AGTATATTCAG AACACACCTT TCCAAGGTAT GTATGCTCTG TTGTCTCTGT
2451 CCTGTCTTCA CTGAGCGCAG GGCTGGAGGC CTCTTAGACA TTCTCTTGG
2501 TCTCTGTTCA GCTGCCCACT GTAGTATCCA CAGTGCCCGA GTTCTCGCTG
2551 TTTTGGCAA TTAACCTCC TTCTACTGCT TTTAGACTAC ACTTACAACA
2601 AGGAAATGTC CCCTCGTGTG ACCATAGATT GAGATTATTA CCACATACCA
2651 CACATAGCCA CAGAAACATC ATCTTGAAAT AAAGAAGAGT TTTGGACAAA
2701 AAAAAAAAAA AAAA
```

## BLAST Results

Entry AF015416 from database EMBL:  
Homo sapiens chromosome 11 from 11p15.5 region, complete sequence.  
Score = 3356, P = 2.0e-144, identities = 672/673

Entry HS263253 from database EMBL:  
human STS SHGC-15914.  
Score = 1143, P = 9.0e-46, identities = 245/255

## Medline entries

No Medline entry

## Peptide information for frame 2

ORF from 326 bp to 1936 bp; peptide length: 537  
Category: similarity to unknown protein

```

1 MEPRVAEAV ETGEEDVIME ALRSYNQEHQ QSFTFDDAQQ EDRKRLAELL
51 VSVLEQGLPP SHRVIWLQSV RILSRDRNCL DPFTSRQSLQ ALACYADISV
101 SEGSVPEASD MDVVLESKLC LCNLVLSPPV AQMLAAEARL VVKLTERVGL
151 YRERSFPHDV QFFDLRLFL LTALRTDVRQ QLFQELKGVR LLTDTLELTL
201 GVTPEGNPPP TLLPSQETER AMEILKVLFN ITLDSIKGEV DEEDAALYRH
251 LGTLLRHCVI IATAGDRTEE FHGHAVNLLG NLPLKCLDVL LTLEPHGDST
301 EFMGVNMDVI RALLIFLEKR LHKTHRLKES VAPVLSVLTE CARMHRPARK
351 FLKAQGWPPP QVLPPLRDVR TRPEVGEMLR NKLVRMLMTHL DTDVKRVAE
401 FLFVLCSSEV PRFIKYTYG NAAGLLAARG LMAGGRPEGQ YSEDEDTDT
451 EYKEAKASIN PVTGRVEEKP PNPMEGMTTEE QKEHEAMKLV TMFDKLSRN
501 VIQPMGMSPR GHLTSLQDAM CETMEQQLSS DPDSDDP

```

## BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKF2phut1\_22o2, frame 2

TREMBL:SPBC3E7\_3 gene: "SPBC3E7.03c"; product: "hypothetical protein";  
S.pombe chromosome II cosmid c3E7., N = 1, Score = 112, P = 0.0023

>TREMBL:SPBC3E7\_3 gene: "SPBC3E7.03c"; product: "hypothetical protein";  
S.pombe chromosome II cosmid c3E7.  
Length = 362

## HSPs:

Score = 112 (16.8 bits), Expect = 2.3e-03, P = 2.3e-03  
Identities = 71/289 (24%), Positives = 124/289 (42%)

```

Query: 215 SQETERAM-EILKVLFNITLDSIKGEVDEEDAALYRHLGTLRLHCVMIATAGDRTEEFHG 273
      SQ+ E + EIL++LF I+ S E DE+ L L+ + +
Sbjct: 12 SQDNEMVLTEILRLFLPISKRSYLKEEDEQKILL-----LVIEIWASSLNNNPNSPLRW 65

Query: 274 HAVN-LLG-NLPLKCLDVLLTLEPHGDSTEFMGVNMDVIRALLIFLEKRLHKTH----RL 327
      HA N LL NL L LD + + T + +I + +LEK L+ +
Sbjct: 66 HATNALLSFNLQLSLDQAIYVSEIACQT----LQSILISREVEYLEKGLNLCFDIAAKY 121

Query: 328 KESVAPVLSVLTECARMHRPARKFLKAQGWPPPQVLPPLRDVRTRP-EVGEMLRNKLVR 386
      + ++ P+L++L + +L P D R + + G+ R L+RL
Sbjct: 122 QNTLPPIAAILLSLFFNIKQNL-----SMLLFPTNDRKQSLQKGSFRCLLLRL 173

Query: 387 MT-HLDTDVKRVAAEFVLVCSSEVPRFIKYTYGNAAGLLAARGLMAGGRPEGQYS--- 442
      +T + + A L LC + + G G A G+ M P + +
Sbjct: 174 LTIPIVEPIGTYIASLLNELCDGDSQQIARIFGAGYAMGISQHSETMPFPSLSKAASPV 233

Query: 443 -EDEDTDDEYKEAKASINPVTGRV--EEKPPNPMEGMTTEEQKEHEAMKLVMTFDKLSRN 499
      + + +E +I+P+TG + +E +++E+KE EA +L +F +L +N
Sbjct: 234 FQKNSRGQENTEENLAIDPITGSMCTNRNKSQRLE-LSQEEKEREERLFYLFQRLEKN 292

```

```
Query:      500 RVIQ 503
           IQ
Sbjct:      293 STIQ 296
```

Pedant information for DKFZphutel\_22o2, frame 2

## Report for DKFZphutel1\_22o2.2

[illegible]

Prosite for DKFZphutel 22o2.2

PS000001	230->234	ASN_GLYCOSYLATION	PDOC000001
PS000005	61->64	PKC_PHOSPHO_SITE	PDOC000005
PS000005	69->72	PKC_PHOSPHO_SITE	PDOC000005
PS000005	84->87	PKC_PHOSPHO_SITE	PDOC000005
PS000005	117->120	PKC_PHOSPHO_SITE	PDOC000005
PS000005	145->148	PKC_PHOSPHO_SITE	PDOC000005
PS000005	218->221	PKC_PHOSPHO_SITE	PDOC000005
PS000005	235->238	PKC_PHOSPHO_SITE	PDOC000005
PS000005	324->327	PKC_PHOSPHO_SITE	PDOC000005
PS000005	463->466	PKC_PHOSPHO_SITE	PDOC000005
PS000005	508->511	PKC_PHOSPHO_SITE	PDOC000005
PS000006	12->16	CK2_PHOSPHO_SITE	PDOC000006
PS000006	34->38	CK2_PHOSPHO_SITE	PDOC000006
PS000006	52->56	CK2_PHOSPHO_SITE	PDOC000006
PS000006	99->103	CK2_PHOSPHO_SITE	PDOC000006
PS000006	104->108	CK2_PHOSPHO_SITE	PDOC000006
PS000006	263->267	CK2_PHOSPHO_SITE	PDOC000006
PS000006	371->375	CK2_PHOSPHO_SITE	PDOC000006

WO 01/12659

PCT/IB00/01496

PS00006	388->392	CK2_PHOSPHO_SITE	PDOC00006
PS00006	442->446	CK2_PHOSPHO_SITE	PDOC00006
PS00006	447->451	CK2_PHOSPHO_SITE	PDOC00006
PS00006	491->495	CK2_PHOSPHO_SITE	PDOC00006
PS00006	515->519	CK2_PHOSPHO_SITE	PDOC00006
PS00006	530->534	CK2_PHOSPHO_SITE	PDOC00006
PS00008	57->63	MYRISTYL	PDOC00008
PS00008	420->426	MYRISTYL	PDOC00008
PS00008	424->430	MYRISTYL	PDOC00008
PS00008	430->436	MYRISTYL	PDOC00008

(No Pfam data available for DKFzphute1\_22o2.2)

DKFZphut1\_23e13

group: metabolism

DKFZphtes3\_15j18 encodes a novel 148 amino acid protein with similarity to 27K heat shock proteins.

The novel protein contains a serine protease of the subtilase family with an aspartic acid-containing active site. Subtilases are an extensive family of serine proteases whose catalytic activity is provided by a charge relay system similar to that of the trypsin family of serine proteases but which evolved by independent convergent evolution. The sequence around the residues involved in the catalytic triad (aspartic acid, serine and histidine) are completely different from that of the analogous residues in the trypsin serine proteases. Thus the novel protein is a new member of this family.

The new protein can find application in modulation of proteinase activity in cells and as a new enzyme for proteomics and biotechnological production processes.

heat shock protein HSP27

strong similarity to heat shock 27K proteins

complete cDNA, complete cds, EST hits

Sequenced by EMBL

Locus: /map="578.9 cR from top of Chr12 linkage group"

Insert length: 1854 bp

Poly A stretch at pos. 1831, polyadenylation signal at pos. 1810

```
1 GGTATTATTA GCTCCTGGCT CCGCTCTAGA CCTCAGCGGT TCTGGCTGCC
51 AGCCTGGGCA GCCTGGGAAG CCTGGGAGGA CGGTGGCTTG CCGGTCTGTC
101 GTGAGGCAGT GCGGACGGGG ACCCTCTGGG ATTCTGCTGG ATCTGCCCGG
151 GGGGTACCTT TTGGGGGCTG GGACCCAGT CGAGGGGACA CAACCGTCCC
201 TGGCAGTGGT TGGTTCTGCT TCTCCCTGCA GAAAAGCAGC ATTTTCGGAA
251 GCTGAAGAAAT AAGCTAGCCC AGCCACACCA CCTTGTGTG TGACCTTGGG
301 CAGGTGGTTC TGTCTCTCTG AGCCTCTGTT TCTCTCTGAG CTGAGCAGCC
351 ACCATGGCTG ACGGTCAGAT GCCCTTCTCC TGCCACTACC CAAGCCGCTT
401 GCGCCGAGAC CCCTTCGGGG ACTCTCCCTT CTCTCTCGC CTGCTGGATG
451 ATGGCTTTGG CATGGACCCC TTCCAGACG ACTTGACAGC CTCTTGGCCC
501 GACTGGGCTC TGCTCTGCTT CTCTCCGCC TGCCAGGCA CCCTAAGGTC
551 GGGCATGGTG CCCCGGGGCC CCACTGCCAC CGCCAGGTTT GGGGTGCCTG
601 CCGAGGGCAG GACCCCCCA CCCTTCCCTG GGGAGCCCTG GAAAGTGTGT
651 GTGAATGTGC ACAGCTTCAA GCCAGAGGAG TTGATGGTGA AGACCAAGA
701 TGGATACGTG GAGGTGCTG GCAACATGA AGAGAAACAG CAAGAAGGTG
751 GCATTGTTTC TAAGAACTTC ACAAGAAAA TCCAGCTTCC TGCAAGGTG
801 GATCCTGTGA CAGTATTTGC CTCACCTTCC CCAGAGGGTC TGCTGATCAT
851 CGAAGCTCCC CAGGTCCCTC CTTACTCAAC ATTTGGAGAG AGCAGTTTCA
901 ACAACGAGCT TCCCCAGGAC AGCCAGGAAG TCACCTGTAC CTGAGATGCC
951 AGTACTGGCC CATCTTGTG TTGTCCCCAA CCCTAGGGCT TCTCTGATTC
1001 CAGGATACAT TACTTTAGCT GAACTCAGAT TTAGTGCAAG TAAATGTTA
1051 GAGGGTGCGG GGGTGAGGAC TGACCACAGA TTCCCTGGAT AGTGTAGTGG
1101 TAGATTTCTC CACAGGATAG CGCAATTGGC AAATCATGCT TGGTTGTGTT
1151 AGGCCAAAAT ACTAGTTTTC CTTTCTTTAC CTTTCTATC TTGATGAAAA
1201 TGTGTGCACAT TCTATAGTTG CAAAACACAT AAAAGGGGAC TTAACATTTC
1251 ACGTTGTATC TTACTTGACG TGAATGCAAG GGTTACTTTT CTCTGGGGAC
1301 CTCCCCCATC ACCCAGGTTT CTACTCTGGG CTCCCGATTC CCATGGCTCC
1351 CAAACCATGC CGCATGGTTT GGTAAATGAA ACCCAGTAGC TAACCCCACT
1401 GTGCTTCCAC ATGCCTGGCC TAAAATGGGT GATATACAGG TCTTATATCC
1451 CCATATGGAA TTTATCCATC AACCACATAA AAACAAACAG TGCCTTCTGC
1501 CCTCTGCCCA GATGTGTCCA GCACGTTCTC AAAGTTTCCA CATTAGCACT
1551 CCCTAAGGAC GCTGGGAGCC TGTCAGTTTA TGATCTGACC TAGGTCCCCC
1601 CTTTCTCTG TCCCTGTGT TTAAGTCGGG ATTTTACAG AGGGAGCTGT
1651 CTCCAGACAG CTCATCAGG AACCAAGCAA AGGCCAGATA GCCTGACAGA
1701 TAGGCTAGTG GTATTGTGTA TATGGGCGGG ACGTGTGTGT CATTATTATT
1751 TGAGTTATGC TGTGTTTAG GGGTAAATAA CAGTAAATAA TTAATAATAA
1801 TAATAATAAT AATAAAGGAG CTGACGTCTT TAAAAAGAA AAAAAAATAA
1851 AAAA
```

## BLAST Results

Entry HS286348 from database EMBL:  
human STS TIGR-A002J47.  
Score = 510, P = 1.2e-16, identities = 102/102

## Medline entries

95394379:  
Cloning and sequencing of a cDNA encoding the canine HSP27 protein.

94110260:  
Physiological and pathological changes in levels of the two  
small stress proteins, HSP27 and alpha B crystallin, in rat  
hindlimb muscles

## Peptide information for frame 3

ORF from 354 bp to 941 bp; peptide length: 196  
Category: strong similarity to known protein  
Prosites motifs: SUBTILASE\_ASP (28-39)

1 MADGQMPFSC HYPSRLRRDP FRDSPLSRL LDDGFGMDPF PDDLTAASWPD  
51 WALPRLSSAW PGTLRSGMVP RGPTATARFG VPAEGRTPPP FPGEPPKVCV  
101 NVHSFKPEEL MVKTKDGYVE VSGKHEEKQQ EGGIVSKNFT KKIQLPAEVD  
151 PVTVFASLSP EGLLIIEAPQ VPPYSTFGES SFNNELPQDS QEVCTCT

## BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFzphutel\_23e13, frame 3

PIR:JC4244 heat-shock 27K protein - dog, N = 1, Score = 304, P =  
4.3e-27

PIR:JN0924 heat shock 27 protein - rat, N = 1, Score = 301, P = 8.9e-27

TREMBL:MM03561\_1 product: "heat shock protein HSP27"; Mus musculus  
heat shock protein HSP27 internal deletion variant b mRNA, complete  
cds., N = 1, Score = 301, P = 8.9e-27

>PIR:JC4244 heat-shock 27K protein - dog  
Length = 209

## HSPs:

Score = 304 (45.6 bits), Expect = 4.3e-27, P = 4.3e-27  
Identities = 80/182 (43%), Positives = 102/182 (56%)

Query: 1 MADGQMPFSC-HYPSRLRRDPFRD-SPLSSRLLDGFGMDPFPPDDLTAASWPDWALPRLSS 58  
M + ++PFS PS DPFRD P SRL D FG+ P++ W W S  
Sbjct: 1 MTERRVPFSLRSPSW---DPFRDWYPAHSRLFDQAFGLPRLPEE----WAQWFG---HS 50

Query: 59 AWPCTLRSGMVP---RGPTATARFGVPAEGR--TPPPFG-----EPWKVCNVVHSF 105  
WPG +R +P GP A A PA R + G + W+V ++V+ F  
Sbjct: 51 GWPGYVRP--IPPAVEGPAAAAAAPPAYSRLSRQLSSGVSEIRQTADRWRVSLDVNHF 108

Query: 106 KPEELMVKTKDGYVEVSGKHEEKQQEGGIVSKNFTKKIQLPAEVD PVTVFASLSPEGLLI 165  
PEEL VKTKDG VE+GKHEE+Q E G +S+ T K LP VDP V +SLSPEG L  
Sbjct: 109 APEELTVKTKDGVVEITGKHEERQDEHGYISRRLTPKYTLPPGVDP TLVSSSLSPEGTLT 168

Query: 166 IEAPQVPPYSTFGE 179  
+EAP P + E  
Sbjct: 169 VEAPMPKPATQSAE 182

## Pedant information for DKFzphutel\_23e13, frame 3

## Report for DKFzphutel\_23e13.3

[LENGTH] 196  
[MW] 21604.37

```

[pI]          5.00
[HOMOL]       PIR:JC4244 heat-shock 27K protein - dog 3e-22
[BLOCKS]      BL01031C
[PIRKW]       blocked amino end 1e-13
[PIRKW]       acetylated amino end 4e-13
[PIRKW]       phosphoprotein 7e-21
[PIRKW]       glycoprotein 2e-11
[PIRKW]       heat shock 7e-21
[PIRKW]       molecular chaperone 4e-13
[PIRKW]       alternative splicing 1e-19
[PIRKW]       eye lens 6e-14
[PIRKW]       stress-induced protein 7e-21
[SUPFAM]      alpha-crystallin 7e-21
[PROSITE]     SUBTILASE_ASP 1
[PROSITE]     MYRISTYL 2
[PROSITE]     CK2_PHOSPHO_SITE 2
[PROSITE]     PKC_PHOSPHO_SITE 6
[PROSITE]     ASN_GLYCOSYLATION 1
[PFAM]        Heat shock hsp20 proteins
[KW]          All_Beta
[KW]          LOW_COMPLEXITY 7.14 %

```

```

SEQ  MADGQMPFSCHYPSRLRRDPFRDSPLSSRLDDGFGMDPFPDDLTA SWPDWALPRLSSAW
SEG  .....xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx
PRD  ccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc

SEQ  PGTLRSGMVPRGPTATARFGVPAEGRTPPPFPGEPWKVCNVVHSFKPEELMVKT KDGYVE
SEG  .....
PRD  cccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc

SEQ  VSGKHEEKQEGGIVSKNFTKKIQLPAEVDPTVTFASLSPEGLLI IEAPQVPPYSTFGES
SEG  .....
PRD  eccchhhhhccccccccccccccccccccccccccccccccccccccccccccccccccc

SEQ  SFNNELPQDSQEVCTT
SEG  .....
PRD  ccccccccccccccccc

```

## Prosites for DKFZphut1\_23e13.3

PS00001	138->142	ASN_GLYCOSYLATION	PDOC00001
PS00005	27->30	PKC_PHOSPHO_SITE	PDOC00005
PS00005	63->66	PKC_PHOSPHO_SITE	PDOC00005
PS00005	76->79	PKC_PHOSPHO_SITE	PDOC00005
PS00005	104->107	PKC_PHOSPHO_SITE	PDOC00005
PS00005	122->125	PKC_PHOSPHO_SITE	PDOC00005
PS00005	140->143	PKC_PHOSPHO_SITE	PDOC00005
PS00006	47->51	CK2_PHOSPHO_SITE	PDOC00006
PS00006	176->180	CK2_PHOSPHO_SITE	PDOC00006
PS00008	62->68	MYRISTYL	PDOC00008
PS00008	132->138	MYRISTYL	PDOC00008
PS00136	28->39	SUBTILASE_ASP	PDOC00125

## Pfam for DKFZphut1\_23e13.3

```

HMM_NAME      Heat shock hsp20 proteins
HMM            *AMMrpPwDWRE....DpDHFeVrMDMPGFKPEEIKvkVEDNNVLvIeG
               A   P++ R   + ++V++++ FKPEE+ VK+ D+ +++++G
Query         77 ARFGVPAEGR-TPPPFPGEPWKVCNVVHSFKPEELMVKT KDG-YVEVSG 123

HMM            EHEREEEREDdkWWHERIYRHFMRFRrLPENVDPdQIkAsMSdNGVLTi
               +HE E++   + + ++ F +++LP +VDP + AS+S++G+L I
Query         124 KHE---EKQQ---EGGIVSKNFTKKIQLPAEVDPTVTFASLSPEGLLI 166

HMM            TVPKpEP*
               ++P ++P
Query         167 EAPQVPP 173

```

DKFZphute1\_23g11

group: uterus derived

DKFZphute1\_23g11 encodes a novel 256 amino acid protein with similarity to *S.pombe* SPAC31G5.12c and *S. cerevisiae* Maf1p.

No informative BLAST results; No predictive prosite, pfam or SCOP motife.

The new protein can find application in studying the expression profile of uterus-specific genes.

similarity to SPAC31G5.12c and Maf1p

complete cDNA, complete cds, EST hits

Sequenced by EMBL

Locus: unknown

Insert length: 1674 bp

Poly A stretch at pos. 1664, polyadenylation signal at pos. 1644

```
1 GGGGGAGGCG GAGGTCGCTC GCTCGCTCGC TCGGCTCGCT GACTCGCCGG
51 AGCGCTCTGT GCGGTCGCGC GGCAGGTCGG TCGCGAGAGC GGGCTCTGTG
101 GAAGGGGGCG AGGCTATGTC GCGGTGGCAG CCCGATGGG CCGGCAGGGC
151 CGGGAGTAAC GGGACGTCGC CGCGGAGCTT CTCCCCCGG ATACAGTGCG
201 GCGCGAGCGG AGGCCGCGGC GCGGCCCTCC GATCTTGAAG AGCCCCGCGT
251 GCGCGGAGCC CGCCCCCGCC TCGGCACCGG CACCGACGCG GAGCGACCCG
301 CCCAGCCAGA CCCGGCCCGG CGCGGCTGA TCTAACCCAG CCAGGCAGGC
351 AATACTAGCC CCTCTGGAGC ACGGAGCTCC TTCCCCAAG ACATGAAGCT
401 ATGGGAGAAC TCGAGCTTTG AAGCCATCAA CTCACAGCTG ACTGTGGAGA
451 CCGGAGATGC CCACATCATT GGCAGGATTG AGAGTACTC ATGTAAGATG
501 GCAGGAGACG ACAACACAT GTTCAAGCAG TTCTGCCAGG AGGGCCAGCC
551 CCACGTGCTG GAGGCACTTT CTCCACCCCA GACTTCAGGA CTGAGCCCCA
601 GCAGACTCAG CAAAAGCCAA GCGGTGAGG AGGAGGGCCC CCTCAGTGAC
651 AAGTGCAGCC GCAAGACCCT CTTCTACCTG ATTGCCACGC TCAATGAGTC
701 CTTCAGGCCCT GACTATGACT TCAGCACAGC CCGCAGCCAT GAGTTCAGCC
751 GGGAGGCCAG CCTTAGCTGG GTGGTGAATG CAGTCAACTG CAGTCTGTTC
801 TCAGCTGTGC GGGAGGACTT CAAGGATCTG AAACCACAGC TGTGGAACGC
851 GGTGGACGAG GAGATCTGCC TGGCTGAATG TGACATCTAC AGCTATAACC
901 CAGACTTGGA CTCAGATCCC TTCGGGGAGG ATGGTAGCCT CTGGTCCTTC
951 AAGTACTTCT TCTACAACAA GCGGCTCAAG CGAATCGTCT TCTTTAGCTG
1001 CCGTTCCATC AGTGGCTCCA CCTACACACC CTCAGAGGCA GGCAACGAGC
1051 TGGACATGGA GCTGGGGGAG GAGGAGGTGG AGGAAGAAAG CAGAAGCAGG
1101 GGCAGTGGGG CCGAGGAGAC CAGCACCATG GAGGAGGACA GGGTCCAGT
1151 GATCTGTATT TGATGAGGAG GAGCCGAGGC CCCAGCTTCA TCCAGCTTCA
1201 ACCAATGCCT GGACCTGTCC ACCTGAGAGG CCCCTGGGGC CTCCCCAGCT
1251 GCTGGGCAGA CCCTGGCGCT GCCACAGTCC TGGCACTGCC CAAGGCCATA
1301 CCTGCTAGC CTTTGGCTC CATCCTGTGG ATGCCCACTC ACCCTCAGA
1351 CTCTGCTGC CCAATGCTGTG GCGGACTTGT TCAGCAGGGG GCCTGGTGGG
1401 AGGAGGAGCT GCCCTGCCCA AATGAATGCG CACAGCAGGG ACAGCTGGAC
1451 CGCAGAGTTT ATTTTGTAT TTCTACTGGG CCTGCACACT CCAGCCAAA
1501 GGTCTGTGG CCGGAGGCC CACGAGCAG CCCAGCAGT CACCGGCTCT
1551 GGTCTTGGG CGGCCCCGGT GCCACCTGT ACCCCACCT CGCCCATTTG
1601 GCCGCTGCA CTGAGTGCA CTTTGCTGCA GCTCGTTTCT TTCCAATAAA
1651 AGTTTCTGTG ACTTAAAAAA AAAA
```

## BLAST Results

No BLAST result

## Medline entries

No Medline entry

## Peptide information for frame 3

ORF from 393 bp to 1160 bp; peptide length: 256  
Category: similarity to known protein



```

1 MKLLENSFE AINSQTLVET GDAHIIGRIE SYSCKMAGDD KHMFKQFCQE
51 GQPHVLEALS PPQTSGLSPLS RLSKSQGGEE EGPLSDKCSR KTLFYLIATL
101 NESFRPDYDF STARSHEFSR EPSLSWVVNA VNCSLFSAVR EDFKDLKPQL
151 WNAVDEEICL AECDIYSYNP DLSDPFGED GSLWSFNFF YNKRLKRIVF
201 FSCRSISGST YTPSEAGNEL DMELGEEVEE EESRSGSGA EETSTMEEDR
251 VPVICI

```

## BLASTP hits

Entry SPAC31G5.12 from database TREMBL:  
 gene: "SPAC31G5.12c"; product: "hypothetical protein"; S.pombe  
 chromosome I cosmid c31G5.  
 Score = 272, P = 9.3e-24, identities = 51/127, positives = 80/127

Entry SPD656.1 from database TREMBL:  
 product: "ORF N150"; Yeast DNA for bfr2+ protein/pad1+ protein/sks1+  
 protein, ORF N133, ORF N150, complete cds, and for ORF N118, partial  
 cds.  
 Score = 263, P = 8.4e-23, identities = 50/127, positives = 79/127

Entry S50986 from database PIR:  
 MAF1 protein - yeast (Saccharomyces cerevisiae) >SWISSPROT:MAF1\_YEAST  
 MAF1 PROTEIN. >TREMBL:SC19492.1 gene: "MAF1"; product: "Maf1p";  
 Saccharomyces cerevisiae Maf1p (MAF1) gene, complete cds.  
 >TREMBL:SC8119.11 gene: "MAF1p"; product: "Maf1p"; S.cerevisiae  
 chromosome IV Cosmid 8119.  
 Score = 180, P = 2.3e-17, identities = 43/133, positives = 75/133

Entry AF098499.2 from database TREMBL:  
 gene: "C43H8.2"; Caenorhabditis elegans cosmid C43H8.  
 Score = 263, P = 9.2e-23, identities = 78/252, positives = 118/252

Alert BLASTP hits for DKFZphut1\_23g11, frame 3

No Alert BLASTP hits found

Pedant information for DKFZphut1\_23g11, frame 3

## Report for DKFZphut1\_23g11.3

```

[LENGTH]      256
[MW]           28869.95
[pI]           4.51
[HOMOL]        TREMBL:SPAC31G5.12 gene: "SPAC31G5.12c"; product: "hypothetical protein";
S.pombe chromosome I cosmid c31G5. 4e-23
[FUNCAT]       06.04 protein targeting, sorting and translocation [S. cerevisiae, YDR005c]
6e-13
[PROSITE]      MYRISTYL      3
[PROSITE]      CK2_PHOSPHO_SITE      5
[PROSITE]      PKC_PHOSPHO_SITE      6
[PROSITE]      ASN_GLYCOSYLATION      3
[KW]           All_Alpha
[KW]           LOW_COMPLEXITY      7.81 %

```

```

SEQ  MKLLENSFEAINSQTLVETGDAHIIGRIEYSCKMAGDDKHMFKQFCQEGQPHVLEALS
SEG  .....
PRD  cccccchhhhhhhhhhhccccceeeccchhhhhccchhhhhhhhhccccceeeccc

SEQ  PPQTSGLSPLSRLSKSQGGEEEGPLSDKCSRKTLFYLIATLNESEFRPDYDFSTARSHEFSR
SEG  .....
PRD  cccccccccccccccccccccccccchhhhhhhhhhhcccccccccccccccccccc

SEQ  EPSLSWVVNAVNCSLFSVREDFKDLKPQLWNAVDEEICLAECDIYSYNPDLDSDPFGED
SEG  .....
PRD  cccccchhhhhhhhhhhhhchhhhhhhhhhhhhhhhhcccccecccccccccccccc

SEQ  GSLWSFNFFYNKRLKRIVFFSCRSISGSTYTPSEAGNELDMELGEEVEEESRSGSGA
SEG  .....
PRD  ccceeeceechhhhhhhhhhhccccccccccccccccchhhhhhhhhhhcccccccc

SEQ  EETSTMEEDRVPVICI
SEG  xx.....
PRD  cccccccccceeeccc

```

## Prosites for DKFZphut1\_23g11.3

PS00001	6->10	ASN_GLYCOSYLATION	PDOC00001
PS00001	101->105	ASN_GLYCOSYLATION	PDOC00001
PS00001	132->136	ASN_GLYCOSYLATION	PDOC00001
PS00005	33->36	PKC_PHOSPHO_SITE	PDOC00005
PS00005	85->88	PKC_PHOSPHO_SITE	PDOC00005
PS00005	89->92	PKC_PHOSPHO_SITE	PDOC00005
PS00005	103->106	PKC_PHOSPHO_SITE	PDOC00005
PS00005	112->115	PKC_PHOSPHO_SITE	PDOC00005
PS00005	202->205	PKC_PHOSPHO_SITE	PDOC00005
PS00006	7->11	CK2_PHOSPHO_SITE	PDOC00006
PS00006	99->103	CK2_PHOSPHO_SITE	PDOC00006
PS00006	212->216	CK2_PHOSPHO_SITE	PDOC00006
PS00006	238->242	CK2_PHOSPHO_SITE	PDOC00006
PS00006	244->248	CK2_PHOSPHO_SITE	PDOC00006
PS00008	66->72	MYRISTYL	PDOC00008
PS00008	181->187	MYRISTYL	PDOC00008
PS00008	239->245	MYRISTYL	PDOC00008

(No Pfam data available for DKFZphut1\_23g11.3)

DKFZphutel\_24c19

group: transmembrane protein

DKFZphutel\_24c19 encodes a novel 195 amino acid protein without similarity to known proteins.

The novel protein contains 1 transmembrane region.  
No informative BLAST results; No predictive prosite, pfam or SCOP motife.

The new protein can find application in studying the expression profile of uterus-specific genes and as a new marker for uterine cells.

unknown  
membrane regions: 1  
Summary DKFZphutel\_24c19 encodes a novel 195 amino acid protein, with  
no similarity to known proteins.

unknown

complete cDNA, complete cds, EST hits  
TRANSMEMBRANE 1

Sequenced by Qiagen

Locus: unknown

Insert length: 769 bp  
Poly A stretch at pos. 746, polyadenylation signal at pos. 735

```
1 ACGAGTCAGC CAAAGATGGC TGC GCCCAGG TAATTGAGC AAAGGCCACA
51 GTGAACCTCCG GCGTGGCTGA GGAAGACCGG AGGAGGCACC CACAGGCTGC
101 TGGGAGGAGA GCATAAGGCT CAAAATGGAA AATCATAAAT CCAATAATAA
151 GGAAAAACATA ACAATTGTTG ATATATCCAG AAAAATTAAC CAGCTTCCAG
201 AAGCAGAAAG GAATCTACTT GAAAATGGAT CGGTTTATGT TGGATTAAAT
251 GCTGCTCTTT GTGGCCTCAT AGCAAAACAGT CTTTTTCGAC GCATCTTGAA
301 TGTGACAAAG GCTCGCATAG CTGCTGGCTT ACCAATGGCA GGGATACCTT
351 TTCTTACAAC AGACTTAACT TACAGATGTT TTGTAAGTTT TCCTTTGAAT
401 ACAGGTGATT TGGATTGTGA AACCTGTACC ATAACACGGA GTGGACTGAC
451 TGGTCTTGTT ATTGGTGGTC TATACCCTGT TTTCTTGGCT ATACCTGTAA
501 ATGGTGGTCT AGCAGCCAGG TATCAATCAG CTCTGTACC ACACAAAGGG
551 AACATCTTAA GTTACTGGAT TAGAACTTCT AAGCCTGTCT TTAGAAAGAT
601 GTTATTTCCT ATTTGCTCC AGACTATGTT TTCAGCATAC CTTGGGTCTG
651 AACAAATATA ACTACTTATA AAGGCCCTTC AGTTATCTGA ACCTGGCAAA
701 GAAATTCACG GATTTTAAAC AAATATGTAA ACAAATAATA AATGGTAAAA
751 ACAAAAAAAA AAAAAAAA
```

## BLAST Results

No BLAST result

## Medline entries

No Medline entry

## Peptide information for frame 2

ORF from 125 bp to 709 bp; peptide length: 195  
Category: putative protein

```
1 MENHKSNNKE NITIVDISRK INQLPEAERN LLENGSVYVG LNAALCGLIA
51 NSLFRRILNV TKARIAAGLP MAGIPFLTDD LTYRCFVSFP LNTGDLDCET
101 CTITRSGLTG LVIGGLYPVF LAIPVNGGLA ARYQSALLPH KGNILSYWIR
151 TSKPVFRKML FPILLQTMFS AYLGSEQYKL LIKALQLSEP GKEIH
```

## BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphut1\_24c19, frame 2

No Alert BLASTP hits found

Pedant information for DKFZphut1\_24c19, frame 2

-----  
Report for DKFZphut1\_24c19.2

```
[LENGTH]      195
[MW]           21527.45
[pI]           9.36
[PROSITE]      MYRISTYL      6
[PROSITE]      CK2_PHOSPHO_SITE      1
[PROSITE]      PKC_PHOSPHO_SITE      3
[PROSITE]      ASN_GLYCOSYLATION      3
[KW]           TRANSMEMBRANE 1

SEQ  MENHKSNNKENITIVDISRKINQLPEAERNLLENGSVYVGLNAALCGLIANSLFRRLNV
PRD  ccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc
MEM  .....

SEQ  TKARIAAGLPAGIPFLTTLTYRCFVSFPLNTGDLDCETCTITRSGTLGLVIGGLYPVF
PRD  hhhhhhhcccccccccccccccccccccccccccccccccccccccccccccccccc
MEM  .....MMMMMMMMMMMMMM

SEQ  LAIPVNGGLAARYQSALLPHKGNILSYWIRTSKPVFRKMLFPILLQTMFSAYLGSEYKYL
PRD  ecccccccccccccccccccccccccccccccccccccccccccccccccccccccccc
MEM  MMM.....

SEQ  LIKALQLSEPGKEIH
PRD  hhhhhhcccccccc
MEM  .....
```

Prosite for DKFZphut1\_24c19.2

PS00001	11->15	ASN_GLYCOSYLATION	PDOC00001
PS00001	34->38	ASN_GLYCOSYLATION	PDOC00001
PS00001	59->63	ASN_GLYCOSYLATION	PDOC00001
PS00005	18->21	PKC_PHOSPHO_SITE	PDOC00005
PS00005	82->85	PKC_PHOSPHO_SITE	PDOC00005
PS00005	151->154	PKC_PHOSPHO_SITE	PDOC00005
PS00006	13->17	CK2_PHOSPHO_SITE	PDOC00006
PS00008	40->46	MYRISTYL	PDOC00008
PS00008	47->53	MYRISTYL	PDOC00008
PS00008	68->74	MYRISTYL	PDOC00008
PS00008	110->116	MYRISTYL	PDOC00008
PS00008	127->133	MYRISTYL	PDOC00008
PS00008	142->148	MYRISTYL	PDOC00008

(No Pfam data available for DKFZphut1\_24c19.2)

DKFZphutell\_24e11

group: intracellular transport and trafficking

DKFZphutell\_24e11 encodes a novel 226 amino acid protein, with similarity to human/mouse golgi 4-transmembrane spanning transporter MTP. MTP may function in the transport of nucleosides and/or nucleoside derivatives between the cytosol and the lumen of an intracellular membrane-bound compartment. Thus, the novel protein also seems to be involved in nucleotide sugar transport.

The new protein can find application in modulating the transport of nucleosides and/or nucleoside derivatives between the cytosol and the lumen of an intracellular membrane-bound compartments.

similarity to 4-TRANSMEMBRANE SPANNING TRANSPORTER MTP

complete cDNA, complete cds, EST hits  
potential start at 184,  
TRANSMEMBRANE 4  
function in the transport of nucleosides and/or nucleoside derivatives  
between the cytosol and  
the lumen of an intracellular membrane-bound compartment?

Sequenced by Qiagen

Locus: /map="8"

Insert length: 2005 bp

Poly A stretch at pos. 1988, polyadenylation signal at pos. 1963

```
1  ACGCGTCCGG  CAGAAGCTCG  GAGCTCTCGG  GGTATCGAGG  AGGCAGGCC
51  GCGGGCGCAC  GGGCGAGCGG  GCCGGGAGCC  GGAGCGGCGG  AGGAGCCGGC
101  ACGACGGCGG  CGGCGGGCTC  CAGGCGAGGC  GGTGACGCT  CCTGAAACT
151  TCGCGGGCGG  CTCGCGCCAC  TCGCGCCGGA  GCGATGAAGA  TGTCGCGGCC
201  CTGGACGCGG  TTCTACTCCA  ACAGCTGCTG  CTTGTGCTGC  CATGTCCGCA
251  CCGGCACCAT  CCTGCTCGGC  GTCTGGTATC  TGATCATCAA  TGCTGTGGTA
301  CTGTTGATTT  TATTGAGTGC  CCTGGCTGAT  CCGGATCAGT  ATAACTTTTC
351  AAGTTCTGAA  CTGGGAGGTG  ACTTTGAGTT  CATGGATGAT  GCCAACATGT
401  GCATTGCCAT  TCGGATTTCT  CTTCTCATGA  TCCTGATATG  TGCTATGGCT
451  ACTTACGGAG  CGTACAAGCA  ACGCGCAGCC  TGGATCATCC  CATTCTCTCG
501  TTACCAGATC  TTTGACTTTG  CCCTGAACAT  GTTGGTTGCA  ATCACTGTGC
551  TTTATTTATCC  AAATCCATT  CAGGAATACA  TACGGCAACT  GCCTCCTAAT
601  TTTCCCTACA  GAGATGATGT  CATGTCAGTG  AATCCTACCT  GTTTGGTCTT
651  TATTATTCTT  CTGTTTATTA  GCATTATCTT  GACTTTTAAG  GGTACTTTGA
701  TTAGCTGTGT  TTGGAAGTGC  TACCGATACA  TCAATGGTAG  GAACTCCTCT
751  GATGTCTCTG  TTTATGTTAC  CAGCAATGAC  ACTACGGTGC  TGCTACCCCC
801  GTATGATGAT  GCCACTGTGA  ATGGTGCTGC  CAAGGAGCCA  CCGCCACCTT
851  ACGTCTCTGC  CTAAGCCTTC  AAGTGGGCGG  AGCTGAGGGC  AGCAGCTTGA
901  CTTTGCAGAC  ATCTGAGCAA  TAGTTCTGTT  ATTTCACTTT  TGCCATGAGC
951  CTCTCTGAGC  TTGTTTGTG  CTGAATGCT  ACTTTTAA  ATTTAGATGT
1001  TAGATTGAAA  ACTGTAGTTT  TCAACATATG  CTTTGCTAGA  ACACGTGTAT
1051  AGATTAACTG  TAGAATCTT  CCTGTACGAT  TGGGGATATA  ACGGGCTTCA
1101  CTAACCTTCC  CTAGGCATTG  AAATCTCCC  CAAATCTGAT  GGACCTAGAA
1151  GTCTGCTTTT  GTACCTGCTG  GGCCCCAAG  TTGGGCATTT  TTCTCTCTGT
1201  TCCCTCTCTT  TTGAAAATGT  AAAATAAAC  CAAAATAGA  CAACTTTTC
1251  TTCAGCCATT  CCAGCATAGA  GAACAAACC  TTATGGAAAC  AGGAATGTCA
1301  ATTGTGTAAT  CATTGTTCTA  ATTAGGTAAA  TAGAAGTCCT  TATGTATGTG
1351  TTACAAGAA  TTCCCCACA  ACATCCTTTA  TGAATGAAGT  TCAATGACAG
1401  TTTGTGTTTG  GTGGTAAAG  ATTTCTCCA  TGGCTGAAT  TAAGACCAT
1451  AGAAAGCAC  AGGCGTGGG  AGCAGTGACC  ATCTACTGAC  TGTCTTGTG
1501  GATCTTGTGT  CCAGGGACAT  GGGGTGACAT  GCCTCGTATG  TGTAGAGGG
1551  TGGAAATGG  GTGTTTGGCG  CTGCATGGGA  TCTGGTGCCC  CTCTTCTCT
1601  GGAATTCAT  CCCCACCCAG  GGCCCGCTT  TACTAAGTGT  TCTGCCCTAG
1651  ATTGGTTCAA  GGAGGTCATC  CACTGACTT  TATCAAGTGG  AATTGGGATA
1701  TATTGTATAT  ACTTCTGCT  AACACATGG  AAAAGGGTTT  TCTTTTCCCT
1751  GCAAGCTACA  TCCTACTGCT  TTGAACCTCC  AAGTATGTCT  AGTCACCTTT
1801  TAAAATGTAA  ACATTTTCAG  AAAAATGAGG  ATTGCCTTCC  TTGTATGGCG
1851  TTTTACCTT  GACTACCTGA  ATTGCAAGG  ATTTTATAT  ATTCATATGT
1901  TACAAAGTCA  GCAACTCTCC  TGTGGTTCA  TTATTGAATG  TGCTGTAAT
1951  TAAGTCGTTT  GCAATTAATA  CAAGTTTGC  CCACATCCAA  AAAAAAAAA
2001  AAAAA
```

## BLAST Results

Entry HS012351 from database EMBL:

human STS SHGC-31823.  
Score = 1629, P = 3.1e-67, identities = 343/354

## Medline entries

96199248:  
Identification of a novel membrane transporter  
associated with intracellular membranes by  
phenotypic complementation in the yeast  
*Saccharomyces cerevisiae*.

## Peptide information for frame 1

ORF from 184 bp to 861 bp; peptide length: 226  
Category: strong similarity to known protein

1 MKMVAPWTRF YSNSCCLCCH VRTGTILLGV WYLIINAVVL LILLSALADP  
51 DQYNFSSSEL GGDFFEMDDA NMCIAIAISL LMILICAMAT YGAYKQRAAW  
101 IIPFFCYQIF DFALNMLVAI TVLIYPNSIQ EYIRQLPPNF PYRDDVMSVN  
151 PTCLVLIILL FISIILTFKG YLISCVWNCY RYINGRNSSD VLVYVTSNDT  
201 TVLLPPYDDA TVNGAAKEPP PPYVSA

## BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphut1\_24e11, frame 1

SWISSPROT:MTRP\_HUMAN GOLGI 4-TRANSMEMBRANE SPANNING TRANSPORTER MTP  
(KIAA0108)., N = 1, Score = 551, P = 2.9e-53

SWISSPROT:MTRP\_MOUSE GOLGI 4-TRANSMEMBRANE SPANNING TRANSPORTER MTP., N  
= 1, Score = 539, P = 5.3e-52

TREMBL:HS304981\_1 product: "E3 protein"; Human retinoic acid-inducible  
E3 protein mRNA, complete cds., N = 1, Score = 127, P = 3.4e-06

>SWISSPROT:MTRP\_HUMAN GOLGI 4-TRANSMEMBRANE SPANNING TRANSPORTER MTP  
(KIAA0108).  
Length = 233

## HSPs:

Score = 551 (82.7 bits), Expect = 2.9e-53, P = 2.9e-53  
Identities = 102/221 (46%), Positives = 148/221 (66%)

Query: 9 RFYSNSCCLCCHVRTGTILLGVWYLIINAVVLLILLSALADPDQY---NFSSSELGGDF- 64  
RFYS CC CCHVRTGTI+LG WY+++N ++ ++L + P+ N +G +  
Sbjct: 13 RFYSTRCCGCCCHVRTGTIILGTWYVMVNLMAILLTVEVTHPNMSPAVNIQYEVIGNYYS 72

Query: 65 -EFMDANMCIAIAISLLMILICAMATYGAYKQRAAWIIPFFCYQIFDFALNMLVAITVL 123  
E M D N C+ A+S+LM +I +M YGA + W+IPFFCY++FDF L+ LVAI+ L  
Sbjct: 73 SERMAD-NACVLFASVLMFIISMLVYGAIYQVGLIPFFCYRLDFVLSCLVAISSL 131

Query: 124 IYPNSIQEYIRQLPPNFYRDDVMSVNPTCLVLIILLFISIILTFKGYLISCVWNCYRYI 183  
Y I+EY+ QLP +FPY+DD++++ +CL+ I+L+F ++ + FK YLI+CVWNCY+YI  
Sbjct: 132 TYLPRIKEYLDQLP-DFPYKDDLLALDSSCLLFIVLVFFALFIIFKAYLINCWNCYKYI 190

Query: 184 NGRNSSDVLVYVTSN-DTIVLLPPYDDATVNGAAKEPPPPYVSA 226  
N RN ++ VY +LP Y+ A V KEPPPPY+ A  
Sbjct: 191 NNRNVPEIAVYPAFEAPPQYVLPYEMA-VKMPKEPPPPYLP 233

## Pedant information for DKFZphut1\_24e11, frame 1

## Report for DKFZphut1\_24e11.1

[LENGTH] 226  
[MW] 25419.11

[pI] 4.65  
 [HOMOL] SWISSPROT:MTRP\_HUMAN GOLGI 4-TRANSMEMBRANE SPANNING TRANSPORTER MTP (KIAA0108).  
 5e-40  
 [PROSITE] CK2\_PHOSPHO\_SITE 3  
 [PROSITE] TYR\_PHOSPHO\_SITE 1  
 [PROSITE] PKC\_PHOSPHO\_SITE 1  
 [PROSITE] ASN\_GLYCOSYLATION 3  
 [KW] SIGNAL PEPTIDE 49  
 [KW] TRANSMEMBRANE 2  
 [KW] LOW\_COMPLEXITY 20.80 %

SEQ MKMVAPWTRFYSNSCLCCHVRTGTILLGVWYLIINAVVLLILLSALADPDQYNFSSEL  
 SEG .....XXXXXXXXXXXXXXXXXXXX  
 PRD ccc  
 MEM .....

SEQ GGDFFMDDANMCIAIAISLLMILICAMATYGAYKQRAAWIIPFCYQIFDFALNMLVAI  
 SEG .....XXXXXXXXXXXXXXXXXXXX  
 PRD ccc  
 MEM .....MM

SEQ TVLIYPNSIQEYIRQLPPNFPYRDDVMSVNPTCLVLIILLFISIIILTFKGYLISCVWNCY  
 SEG .....XXXXXXXXXXXXXXXXXXXX  
 PRD hhhcc  
 MEM MMMMM.....MM

SEQ RYINGRNSSDVLVYVTSNDTTVLPPYDDATVNGAAKEPPPPYVSA  
 SEG .....  
 PRD ecc  
 MEM .....

#### Prosites for DKFZphut1\_24e11.1

PS00001	54->58	ASN_GLYCOSYLATION	PDOC00001
PS00001	187->191	ASN_GLYCOSYLATION	PDOC00001
PS00001	198->202	ASN_GLYCOSYLATION	PDOC00001
PS00005	167->170	PKC_PHOSPHO_SITE	PDOC00005
PS00006	56->60	CK2_PHOSPHO_SITE	PDOC00006
PS00006	128->132	CK2_PHOSPHO_SITE	PDOC00006
PS00006	196->200	CK2_PHOSPHO_SITE	PDOC00006
PS00007	186->195	TYR_PHOSPHO_SITE	PDOC00007

(No Pfam data available for DKFZphut1\_24e11.1)

DKF2phutel\_24j6

-----

group: cell structure and motility

DKF2phutes1\_24j6 encodes a novel 571 amino acid protein with strong similarity to rat cell adhesion regulator (CAR1).

The novel protein is very similar to Car1 and thus seems to be involved in regulation cell-cell adhesion. It contains a RGD cell attachment site.

The new protein can find application in modulation of cell-cell-adhesion.

strong similarity to rat CAR1 A.thaliana T19C21.5

complete cDNA, complete cds, EST hits  
potential frame shift at Bp 1241 according to CAR1  
but frame shift might be in CAR1 sequence!  
ESTs T73366 AA362984 confirm this sequence

Sequenced by Qiagen

Locus: /map="939.9 cR from top of Chr2 linkage group"

Insert length: 3333 bp

Poly A stretch at pos. 3316, no polyadenylation signal found

```
1  ACGCGTCCGA  GCTGGCTCAG  GCGGTCCGCT  AGGCTCGGAC  GACCTGCTGA
51  GCCTCCCAAA  CCGCTTCCAT  AAGGCTTTGC  CTTTCCAAC  TCAGCTACAG
101  TGTGTAGCTAA  GTTTGGAAAG  AAGGAAAAAA  GAAAAATCC  GGGCCCCCTT
151  TCTTTTGTTC  TTTGCCAAAG  TCGTCGTGTG  AGTCTTTTG  CCAAAGGCTG
201  TGTGTTTTT  AGAGGTGCTA  TCTCCAGTTC  CTTGCACTCC  TGTTAACAAG
251  CACCTCAGCG  AGAGCAGCAG  CAGCGATAGC  AGCCGACAG  GAGCCAGCGG
301  GGTCCGCTAG  TGTCTAGACC  AGGGCGGGAG  ATCACAACCG  CCAGAGAGGA
351  TGCTGTGGAT  CCTTGGCCGA  CTACCTGACC  TCTGCAAAAT  TCCTTCTCTA
401  CCTTGGTCAT  TCTCTCTCTA  CTTGGGGAGA  TCGGATGTGG  CACTTTGCGG
451  TGTCTGTGTT  TCTGGTAGAG  CTCTATGGAA  ACAGCCTCCT  TTTGACAGCA
501  GTCTACGGGC  TGGTGGTGGC  AGGGTCTGTT  CTGGTCCTGG  GAGCCATCAT
551  CGGTAGCTGG  GTGGACAAGA  ATGCTAGACT  TAAAGTGGCC  CAGACCTCGC
601  TGGTGGTACA  GAATGTTTCA  GTCATCCTGT  GTGGAATCAT  CCTGATGATG
651  GTTTTCTTAC  ATAAACATGA  GCTTCTGACC  ATGTACCATG  GATGGGTCTT
701  CACTTCTCTG  TATATCCTGA  TCATCACTAT  TGCAATATAT  GCAAAATTTG
751  CCAGTACTGC  TACTGCAATC  ACAATCCAAA  GGGATTGGAT  TGTGTTGTTT
801  GCAGGAGAAG  ACAGAAGCAA  ACTAGCAAA  ATGAATGCCA  CAATACGAAG
851  GATTGACCAG  TTAACCAACA  TCTTAGCCCC  CATGGCTGTT  GGCCAGATTA
901  TGACATTGG  CTCCCAGCTC  ATCGGCTGTG  GCTTTATTTC  GGGATGGAA  C
951  TTGGTATCCA  TGTGCGTGGA  GTACGTCCTG  CTCTGGAAGG  TTTACAGAA
1001  AACCCAGCT  CTAGCTGTGA  AAGCTGGTCT  TAAAGAAGAG  GAAACTGAAT
1051  TGAACAGCT  GAATTTACAC  AAAGATACTG  AGCCAAAACC  CCTGGAGGGA
1101  ACTCATCTAA  TGGGTGTGAA  AGACTCTAAC  ATCCATGAGC  TTGAACATGA
1151  GCAAGAGCCT  ACTTGTGCCT  CCCAGATGGC  TGAGCCCTTC  CGTACCTTCC
1201  GAGATGGATG  GGTCTCCTAC  TACAACCAGC  CTGTGTTTCT  GGCTGGCATG
1251  GGTCTTGCTT  TCCTTTATAT  GACTGTCCCTG  GGCTTTGACT  GCATCACCAC
1301  AGGGTACGCC  TACACTCAGG  GACTGAGTGG  TTCCATCCTC  AGTATTTTGA
1351  TGGGAGCATC  AGCTATAACT  GGAATAATGG  GAACTGTAGC  TTTTACTTGG
1401  CTACGTCGAA  AATGTGGTTT  GGTTCGGACA  GGTCTGATCT  CAGGATTGGC
1451  ACAGCTTTCC  TGTGTGATCT  TGTGTGTGAT  CTCTGTATTC  ATGCCTGGAA
1501  GCCCCTGGA  CTTGTCCGTT  TCTCCTTTTG  AAGATATCCG  ATCAAGGTTT
1551  ATTCAGGAG  AGTCAATTAC  ACCTACCAAG  ATACCTGAAA  TTACAACCTGA
1601  AATATACATG  TCTAATGGGT  CTAATTCTGC  TAATATTGTC  CCGGAGACAA
1651  GTCCGTAATC  TGTGCCCATA  ATCTCTGTCA  GTCTGCTGTT  TGCAGGCGTC
1701  ATTGCTGCTA  GAATCGGTCT  TTGGTCCTTT  GATTTAAGTG  TGACACAGTT
1751  CTTGCAAGAA  AATGTAATTG  AATCTGAAAG  AGGCATTATA  AATGGTGTAC
1801  AGAATCCCAT  GAACTATCTT  CTTGATCTTC  TGCATTTCAT  CATGGTCATC
1851  CTGGCTCCAA  ATCCTGAAGC  TTTTGGCTTG  CTCGATTGTA  TTTTCACTCT
1901  CTTTGTGGCA  ATGGGCCACA  TTATGTATTT  CCGATTGTGC  CAAAATACTC
1951  TGGGAAACAA  GCTCTTTGCT  TGCGGTCCTG  ATGCAAAAGA  AGTTAGGAAG
2001  GAAAAACAAG  CAAATACATC  TGTGTTTGA  GACAGTTTAA  CTGTTGCTAT
2051  CCTGTACTTA  GATTATATAG  AGCACATGTG  CTTATTTTGT  ACTGCAGAA  T
2101  TCCAATAAAT  GGCTGGGTGT  TTTGCTCTGT  TTTTACCACA  GCTGTGCCCT
2151  GAGAACTAAA  AGCTGTTTAG  GAAACCTAAG  TCAGCAGAAA  TTAAGTGATT
2201  AATTTCCCTT  ATGTTGAGGC  ATGGAAAAAA  AATTGGAAAA  GAAAAACTCA
2251  GTTTAAATAC  GGAGACTATA  ATGATAACAC  TGAATCCCTC  TATTTCTCAT
2301  GAGTAGATAC  AATCTTACGT  AAAAGAGTGG  TTAGTCACGT  GAATTCAGTT
2351  ATCATTGTAC  AGATTCTTAT  CTGTACTAGA  ATTCAGATAT  GTCAGTTTTC
2401  TGCAAACTC  ACTCTTGTTC  AAGACTAGCT  AATTTATTTT  TTTGCATCTT
2451  AGTTATTTT  AAAACAAAAT  TCTTCAAGTA  TGAAGACTAA  ATTTTGATAA
2501  CTAATATTAT  CCTATTGAT  CCTATTGATC  TTAAGGTATT  TACATGTATG
```



```

2551 TGGAAAAACA AACACTTAA CTAGAATTCT CTAATAAGGT TTATGGTTTA
2601 GCTTAAAGAG CACCTTTGTA TTTTATTAT CAGATGGGGC AACATATTGT
2651 ATGAAGCATA TGTAAGCACT CACAGCATGG TTATCATGTA AGCTGCAGGT
2701 AGAAGCAAAG CTGTAAAGTA GATTATCAC ACAATGACTG CATACAGACT
2751 TCAAAATATGT CAATAGTTTG GTCATAGAAC CTAGAAGCCA AAAGCCACAC
2801 AGAAGGGCAA GAATCCCAAT TTAACCTCAT TTATCATCAT TAGTGATCTG
2851 TGTGTAGAA CATGAGGGTG TAAGCCTTCA GCCTGGCAAG TTACATGTAG
2901 AAAGCCCACT CTTGTGAAGG TTTTGTTTA CAAATCACTT GATTTAACAC
2951 ACTCAGGTAG AATATTTTAA TTTTACTGT TTTATACCCA GAAGTTATTT
3001 CTACATTGTT CTACAGCAAG AATATTCATA AAAGTATCCC TTCAAATGC
3051 CTTTGAGAAG AATAGAAGAA AAAAAGTTTG TATATATTTT AAAAAATTGT
3101 TTTAAAAGTC AGTTTGCAAC ATGCTGTGAC CAAGATGGTA CTTTGCCTTA
3151 ACCGTTTATA TGCACCTTCA TGGAGACTGC AATACGTTGC TATGAGCACT
3201 TTCCTTTATC TTGGAGTTTA ATCCTTTGCT TCATCTTCT ACAGTATGAC
3251 ATAATGATTT GCTATGTTGT AAAATCTTTG TAAAAAATTT CTATATAAAA
3301 ATATTTTGAA AATCTTAAAA AAAAAAAAAA AAA

```

## BLAST Results

Entry HS389210 from database EMBL:  
human STS SHGC-10164.  
Score = 1592, P = 1.5e-64, identities = 346/364

Entry HS933343 from database EMBL:  
human STS WI-16551.  
Score = 1193, P = 5.7e-46, identities = 241/244

## Medline entries

No Medline entry

## Peptide information for frame 3

ORF from 315 bp to 2027 bp; peptide length: 571  
Category: strong similarity to known protein

```

1 MTRAGDHNRO RGCCGSLADY LTSAKFLLYL GHSLSWTGDR MWHFAVSVFL
51 VELYGNLLLL TAVYGLVVAG SVLVLGAIIG DWVDKNARLK VAQTSLVVQN
101 VSVILCGIIL MMVFLHKHEL LTMVHGWLVT SCYILITIA NIANLASTAT
151 AITTIQRDWIV VVAGEDRSKL ANMNATIRRI DQLTNILAPM AVGQIMTFGS
201 PVIGCGFISG WNLVSMCWEY VLLWKVYQRT PALAVKAGLK EEETELKQLN
251 LHKDTEPKPL EGTLMGVKD SNIHELEHEQ EPTCASQMAE PFRTFRDQWV
301 SYYNQPVFLA GMGLAFLYMT VLGFDCTTG YAYTQGLSGS ILSILMGASA
351 ITGIMGTVAE TWLRRKGLV RTGLISGLAQ LSCLILCVIS VFMPGSPDL
401 SVSPFEDIRS RFIQGESITP TKIPEITTEE YMSNGSNSAN IVPETSPESV
451 PIISVSLLFA GVIAARIGLV SFDLTVTQLL QENVIESERG IINGVQNSMN
501 YLLDLLHFIM VILAPNPEAF GLLVLISVSF VAMGHIMYFR FAQNTLGNKL
551 FACGPDAKEV RKENQANTSV V

```

## BLASTP hits

No BLASTP hits available

## Alert BLASTP hits for DKFZphut1\_24j6, frame 3

TREMBLNEW:U76714\_1 gene: "CAR1"; product: "cell adhesion regulator";  
Rattus norvegicus cell adhesion regulator (CAR1) mRNA, complete cds., N  
= 1, Score = 1472, P = 7.2e-151

TREMBL:AC004683\_5 gene: "T19C21.5"; Arabidopsis thaliana chromosome II  
BAC T19C21 genomic sequence, complete sequence., N = 2, Score = 437, P  
= 2.8e-60

TREMBL:AF039046\_2 gene: "R09B5.4"; Caenorhabditis elegans cosmid  
R09B5., N = 2, Score = 323, P = 1.5e-43

>TREMBLNEW:U76714\_1 gene: "CAR1"; product: "cell adhesion regulator";  
Rattus norvegicus cell adhesion regulator (CAR1) mRNA, complete cds.  
Length = 405

**HSPs :**

Score = 1472 (220.9 bits), Expect = 7.2e-151, P = 7.2e-151  
Identities = 288/319 (90%), Positives = 297/319 (93%)

Query:	1	MTRAGDHNRRQCCGSLADYLTSAKFLLLYLGHSLSLTWGD RMWHFAVSVFLVELYGNLSLL	60
		MTS+ D Q CGCGSLA+YLTSAKFLLLYLGHSLSLTWGD RMWHFAVSVFLVELYGNLSLL	
Sbjct:	1	MYKSRDTHQEGGCGSLANYLTSAKFLLLYLGHSLSLTWGD RMWHFAVSVFLVELYGNLSLL	60
Query:	61	TAVYGLVVGASVVLVGAIGDWVDKNARLKVAQTSLLVQNVSVILCGIILMMVFLHKHEL	120
		TAVYGLVVGASVVLVGAIGDWVDKNARLKVAQTSLLVQNVSVILCGIILMMVFLHK+EL	
Sbjct:	61	TAVYGLVVGASVVLVGAIGDWVDKNARLKVAQTSLLVQNVSVILCGIILMMVFLHKHEL	120
Query:	121	LTMHYHGWLTS CYILIITIANIANLASTATAITIQRDWIVVVAGEDRSKLANMNATIRRI	180
		L MYHGWLVT CYILIITIANIANLASTATAITIQRDWIVVVAGE+RS+LA+MNATIRRI	
Sbjct:	121	LNMYHGWLTVCYILIITIANIANLASTATAITIQRDWIVVVAGENRSRLADMNATIRRI	180
Query:	181	DQLTNILAPMAVGQIMTFGSPVIGCGFISGWNLSVMCEYVLLWKVYQKTPALAVKAGLK	240
		DQLTNILAPMAVGQIMTFGSPVIGCGFISGWNLSVMCEY LLWKVYQKTPALAVKA LK	
Sbjct:	181	DQLTNILAPMAVGQIMTFGSPVIGCGFISGWNLSVMCEYFLLWKVYQKTPALAVKAAKL	240
Query:	241	EEETELKQLNLHKDTEPKPLEGTHLMGVKDSNIHELEHEQEPTCASQMAEPFTRFDGWV	300
		EE+ELKQL KDTPEKPLEGTHLM KDSNI EE QEPTCASQ+AEPFTRFDGWV	
Sbjct:	241	VEESELKQLTSPKDTPEKPLEGTHLMGEKDSNIRELECEQEPTCASQIAEPFTRFDGWV	300
Query:	301	SYYNQPVFLAGMGGF-LY 318	
		SYYNQPVFL G FLY	
Sbjct:	301	SYYNQPVFLGWHGFGGFLY 319	

Pedant information for DKFZphutel 24j6, frame 3

## Report for DKFZphutel1\_24j6.3

```

[LENGTH]          571
[MW]               62542.72
[pI]               6.08
[HOMOL]            TREMBL:U76714_1 gene: "CAR1"; product: "cell adhesion regulator"; Rattus
norvegicus cell adhesion regulator.(CAR1) mRNA, complete cds. 1e-141
[BLOCKS]           BL00341D
[PROSITE]          MYRISTYL . 15
[PROSITE]          MITOCH CARRIER 1
[PROSITE]          CK2_PHOSPHO_SITE 6
[PROSITE]          PROKAR LIPOPROTEIN 1
[PROSITE]          PKC_PHOSPHO_SITE 4
[PROSITE]          ASN_GLYCOSYLATION 4
[PFAM]             Laminin B (Domain IV)
[KW]               TRANSMEMBRANE 4
[KW]               LOW COMPLEXITY 8.76 %

```

```

SEQ      MTRAGDHNRRQGCCGSLADYLTSAKFLLYLGHLSLSTWGRDMWHFAVSVFLVELYGNLSLL
SEG
PRD      cccccccccccccccccchhhhhhhhhheeeccceeeccchhhhhhhhhheeecccccce
MEM      .....MMMMMMMMMMMMMM

SEQ      TAVYGLVVGASVLVLGAIIGDWDKNARLKVAQTSLVVQNVSVILCGIILMMVFLHKHEL
SEG      .xxxxxxxxxxxxxxxx
PRD      ehhhhhhhhhccceeecccccchhhhhhhhhhhheeeccchhhhhhhhhhhhhhhhhhh
MEM      MMMMMMMMMMMMMMMM.....MMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMM

SEQ      LTMYHGWVLTSCYLIITIANIANLASTATAITIQRDWIVVAGEDRSKLANMNATIRRI
SEG      .xxxxxxxxxxxxxxxxxxxx
PRD      hhccccchhhhhhhhhhhhhhhhhhhhhheeeccceeeeeecccccchhhhhhhhhhh
MEM      MMMMMMMM.....

SEQ      DQLTNILAPMAVGQIMTFGSPVIGCGFISGWNLVSMCVEYVLLWKVYQKTPALAVKAGLK
SEG
PRD      hhhhhhhccceeeceeeeeeccceeeeeeccchhhhhhhhhhhhhhhhhccchhhhhhhhh
MEM      .....

SEQ      EEETELKQLNLHKDTEPKPLEGTHLMGVKDSNIHELEHEQEPTCASQMAEPFRTFRDGVV
SEG
PRD      hhhhhhhhhhhccccccccccceeeeeecccccccccccccccccccccccccccccce
MEM      .....

SEQ      SYYNQPVFLAGMGLAFLYMTVLGFDICITTYAYTQGLSGSILSMGASAITGIMGTAVF
SEG
PRD      eeccceeeccchhhhhhhccccceeeeeecccccceeeeeeccceeeeeehehhhhhh

```

```
MEM      .....

SEQ      TWLRRKCGLVRTGLISGLAQLSCLILCVISVFMGPSPLDLSVPFEDIRSRFIQGESITP
SEG      xxxxxxxxxx
PRD      hhhhhhhccccccccchhhhhhhhhhhhhhhhhhhccccccccccccchhhhhhhcccccccc
MEM      .....

SEQ      TKIPEITTEIYMSNGSNSANIVPETSPEVPIISVSLLFAGVIAARIGLWSFDLTVTQLL
SEG      xxxxxxxxxx
PRD      cccccceeeeeeccccccccccccccccccccceeeeee hhhhhhhhhhhccccchhhhhhhhh
MEM      .....MMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMM.....

SEQ      QENVIESERGIINGVQNSMYNLDDLHFIMVILAPNEAFGLLVLSVSFVAMGHIMYFR
SEG      .....
PRD      hhhhhccccceeeeeeccchhhhhhhhhhhheeeccccccccceeeeeeccccccccceeee
MEM      .....MMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMM.....

SEQ      FAQNTLGNKLFACGPDACEVRKENQANTSVV
SEG      .....
PRD      eccccceeeeeeccccchhhhhhhhhhhcccccc
MEM      .....
```

Prosites for DKFZphut1\_24j6.3

PS000001	100->104	ASN_GLYCOSYLATION	PDCC000001
PS000001	174->178	ASN_GLYCOSYLATION	PDCC000001
PS000001	434->438	ASN_GLYCOSYLATION	PDCC000001
PS000001	567->571	ASN_GLYCOSYLATION	PDCC000001
PS000005	23->26	PKC_PHOSPHO_SITE	PDCC000005
PS000005	176->179	PKC_PHOSPHO_SITE	PDCC000005
PS000005	294->297	PKC_PHOSPHO_SITE	PDCC000005
PS000005	487->490	PKC_PHOSPHO_SITE	PDCC000005
PS000006	16->20	CK2_PHOSPHO_SITE	PDCC000006
PS000006	36->40	CK2_PHOSPHO_SITE	PDCC000006
PS000006	294->298	CK2_PHOSPHO_SITE	PDCC000006
PS000006	396->400	CK2_PHOSPHO_SITE	PDCC000006
PS000006	403->407	CK2_PHOSPHO_SITE	PDCC000006
PS000006	445->449	CK2_PHOSPHO_SITE	PDCC000006
PS000008	12->18	MYRISTYL	PDCC000008
PS000008	65->71	MYRISTYL	PDCC000008
PS000008	76->82	MYRISTYL	PDCC000008
PS000008	193->199	MYRISTYL	PDCC000008
PS000008	267->273	MYRISTYL	PDCC000008
PS000008	311->317	MYRISTYL	PDCC000008
PS000008	336->342	MYRISTYL	PDCC000008
PS000008	339->345	MYRISTYL	PDCC000008
PS000008	353->359	MYRISTYL	PDCC000008
PS000008	368->374	MYRISTYL	PDCC000008
PS000008	373->379	MYRISTYL	PDCC000008
PS000008	435->441	MYRISTYL	PDCC000008
PS000008	461->467	MYRISTYL	PDCC000008
PS000008	490->496	MYRISTYL	PDCC000008
PS000008	494->500	MYRISTYL	PDCC000008
PS000113	122->133	PROKAR_LIPOPROTEIN	PDCC000113
PS002115	404->414	MITOCAR_CARRIER	PDCC001189

Pfam for DKFZphutel 24j6.3

HMM_NAME	Laminin B (Domain IV)		
HMM	*YWRIPERFLGDQvTsYGGkLe*		
	Y+R	+ LG+++ + G + +	
Query	538	YFRFAQNTLGKFLFACGPDAK	558

DKF2phut1\_2h3

-----

group: differentiation/development

DKF2phut1\_2h3 encodes a novel 267 amino acid protein, with similarity to ITM2 (integral membrane protein 2) of chicken and mouse.

The novel protein contains a prenyl group binding site (CAAX box) and seems to be post-translationally modified by the attachment of either a farnesyl or a geranyl-geranyl group. The similar gallus G. protein E25 a marker for chondro-osteogenic differentiation.

The new protein can find application as a useful marker for chondro-osteogenic cell differentiation and for the modulation of chondro-osteogenic cell differentiation.

strong similarity to mouse E25 and gallus E3-16

complete cDNA, EST hits  
complete cds according to E25 start at Bp 56  
putative transmembrane protein (1 TM)

Sequenced by AGOWA

Locus: unknown

Insert length: 2033 bp

Poly A stretch at pos. 2007, polyadenylation signal at pos. 1986

```

1 GGACCGAGGC TGCACCGGCA GAGGCTGCGG GCGGACGCG CGGGCCGGCG
51 CAGCCATGGT GAAGATTAGC TTCCAGCCCG CCGTGGCTGG CATCAAGGGC
101 GACAAGGCTG ACAAGGCGTC GCGCTCGGCC CCTGCGCCGG CCTCGGCCAC
151 CGAGATCTCT CTGACGCCGG CTAGGGAGGA GCAGCCCCCA CAACATCGAT
201 CCAAGAGGGG GAGCTCAGTG GCGGCGGTGT GCTACCTGTC GATGGGCATG
251 GTCTGTCTGC TCATGGGCCT CGTGTTCGCC TCTGTCTACA TCTACAGATA
301 CTTCTTTCTT GCACAGCTGG CCGAGATAA CTTCTCCGC TGTGGTGTGC
351 TGTATGAGGA CTCCTGTGCC TCCAGGTCC GGAAGTATG GAGCTGGAA
401 GAGGATGTGA AATCTACCT CGACGAGAAC TACGAGCGCA TCAACGTGCC
451 TGTGCCCCAG TTGGCGGCG GTGACCTGTC AGACATCATC CATGACTTCC
501 AGCGGGGTCT GACTGCGTAC CATGATATCT CCTGGACAA GTGCTATGTC
551 ATCGAACTCA ACACCAACAT TGTGCTGCCC CCTCGCAACT TCTGGGAGCT
601 CCTCATGAAC GTGAAGAGGG GGACCTACCT GCCGCAGACG TACATCATCC
651 AGGAGGAGAT GGTGGTCACG GAGCATGTCA GTGACAAGGA GGCCCTGGGG
701 TCCTTCATCT ACCACCTGTG CAACGGGAAA GACACCTACC GGCTCCGGCG
751 CCGGGCAACG CGGAGGCGGA TCAACAAGCG TGGGGCCAAG AACTGCAATG
801 CCATCCGCCA CTTCGAGAAC ACCTTCCTGG TGGAGACGCT CATCTCGCGG
851 GTGGTGTGAG GCCCTCCTCC CCCAGAACCC CTGCGGTGT TCCTCTTTTC
901 TTTCTTCCAG CTGCTCTCTG GCCCTCCTCC TTCCCTCTGC TTAGCTTTGA
951 CTTTGGACGC GTTCTATAG AGGTGACATG TCTCTCCATT CCTCTCCAAC
1001 CTTGCCCCACC TCCCTGTACC AGAGCTGTGA TCTCTCGGTG GGGGGCCCAT
1051 CTCTGCTGAC CTGGGTGTGG CGGAGGGAGA GGCGATGCTG CAAAGTGTTC
1101 TCTGTGTCCC ACTGTCTTGA AGCTGGGCCT GCCAAAGCCT GGGCCACAG
1151 CTGACCCGGC AGCCCAAGGG GAAGGACCGG TTGGGGGAGC CGGGCATGTG
1201 AGGCCCTGGG CAAGGGGATG GGGCTGTGGG GCGGGGGCGG CATGGGCTTC
1251 AGAAGTATCT GCACAATTAG AAAAGTCTCT AGAAGCTTTT TCTTGGAGGG
1301 TACACTTTCT TCACTGTCCC TATTCTTAGA CCTGGGGCTT GAGCTGAGGA
1351 TGGGACGATG TGCCAGGGA GGGACCCACC AGAGCACAAG AGAAGGTGGC
1401 TACCTGGGGG TGTCCAGGG ACTCTGTGAG TGCTTCAGC CCACCAGCAG
1451 GAGCTTGGAG TTTGGGAGT GGGGATGAGT CCGTCAAGCA CAACTGTTCT
1501 CTGAGTGGAA CCAAGAAGC AAGGAGCTAG GACCCCAAGT CCTGCCCCCC
1551 AGGAGCACAA GCAGGTCCC CTCAGTCAAG GCAGTGGGAT GGGCGGCTGA
1601 GGAACGGGGC AGGCAAGGTC ACTGCTCAGT CACGTCCAGC GGGGACGAGC
1651 CGTGGGTTCT GCTGAGTAGG TGGAGCTCAT TGCTTTCTCC AAGCTTGGAA
1701 CTGTTTTGAA AGATAACACA GAGGGAAAGG GAGAGCCACC TGGTACTTGT
1751 CCACCCTGCC TCCTCTGTTC TGAAATTCCA TCCCCTCAG CTTAGGGGAA
1801 TGCACCTTTT TCCCTTCTCT TCTACTTTT GCATGTTTTT ACTGATCATT
1851 CGATATGCTA ACCGTTCTCA GCCCTGAGCC TTGGAGAGGA GGGCTGTAAC
1901 GCCTTCAGTC AGTCTCTGGG GATGAAACTC TTAATGCTT TGTATATTTT
1951 CTCAATTAGA TCTCTTTTCA GAAGTGTCTA TAGAACATA AAAATCTTTT
2001 ACTTCTGAAA AAAAAAAAAA AAAAGGGCGG CCG
```

## BLAST Results

-----

Entry B64417 from database EMBL:  
CIT-HSP-2023A7.TR CIT-HSP Homo sapiens genomic clone 2023A7.  
Length = 715  
Plus Strand HSPs:

### Medline entries

Peptide information for frame 2

BLASTP hits

Pedant information for DKF2phutel 2h3, frame 2

```

[LENGTH]      267
[MW]           30253.96
[pI]           8.16
[HOMOL]        SWISSNEW:ITMB_CHICK INTEGRAL MEMBRANE PROTEIN 2B (TRANSMEMBRANE PROTEIN E3-16).
               1e-49
[PROSITE]      MYRISTYL      4
[PROSITE]      PRENYLATION   1
[PROSITE]      CAMP_PHOSPHO_SITE 3
[PROSITE]      CK2_PHOSPHO_SITE 3
[PROSITE]      TYR_PHOSPHO_SITE 1
[PROSITE]      PKC_PHOSPHO_SITE 4
[PROSITE]      ASN_GLYCOSYLATION 1
[KW]           TRANSMEMBRANE 1
[KW]           LOW COMPLEXITY 15.36 %

```

```

SEQ      MVKISFQPAVAGIKGDKADKASASAPASATEILLTPAREEQPPQHRSKRGSSVGGVCY
SEG      .....XXXXXXXXXXXXXXXXXXXXX
PRD      cccccccccchhhhhhhhhhhhhhhhhhhccccceeeccccccccccccccccccccchh
MEM      .....MMM

SEQ      LSMGMVLLMGLVFASVYIYRYFFLAQLARDNFFRCGLYEDSLSSQVRTQMELEEDVXI
SEG      .....XXXXXXXXXXXX
PRD      hhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhccccceeeccccccccchhhhhhhhhhhhh
MEM      MMMMMMMMMMMMMMMMMMMMMMMMMMMM

SEQ      YLDENYERINVPVPFGGGDPADIINDFORGLTAYHDISLDKCYVIELNTTIVLPPRNFW
SEG      .....
PRD      hhccccceeeccccccccccccchhhhhhhhhhhhhhhhhccccceeeccceeccccchhh
MEM      .....

SEQ      ELLMNVKRGTYLPQTYIIQEEMVVTEHVSDEKALGSFYIHLNCGKDTYRLRRRATRRRIN
SEG      .....XXXXXXXXXXXXXXXXXXXX
PRD      hhhhhhhccccccccceeeehhhhhhhccccchhhhhheeeccccchhhhhhhhhhhhhhhhh
MEM      .....

SEQ      KRGAKNCNAIRHFENTFVETLICGVV
SEG      xx.....
PRD      hhhhhccccceeeccccchhhhhheeeccc
MEM      .....

```

PS00001	169->173	ASN_GLYCOSYLATION	PDOC00001
PS00004	50->54	CAMP_PHOSPHO_SITE	PDOC00004
PS00004	187->191	CAMP_PHOSPHO_SITE	PDOC00004
PS00004	232->236	CAMP_PHOSPHO_SITE	PDOC00004
PS00005	49->52	PKC_PHOSPHO_SITE	PDOC00005
PS00005	209->212	PKC_PHOSPHO_SITE	PDOC00005
PS00005	227->230	PKC_PHOSPHO_SITE	PDOC00005
PS00005	235->238	PKC_PHOSPHO_SITE	PDOC00005
PS00006	30->34	CK2_PHOSPHO_SITE	PDOC00006
PS00006	110->114	CK2_PHOSPHO_SITE	PDOC00006
PS00006	209->213	CK2_PHOSPHO_SITE	PDOC00006
PS00007	119->127	TYR_PHOSPHO_SITE	PDOC00007
PS00008	52->58	MYRISTYL	PDOC00008
PS00008	71->77	MYRISTYL	PDOC00008
PS00008	138->144	MYRISTYL	PDOC00008
PS00008	243->249	MYRISTYL	PDOC00008
PS00294	264->268	PRENYLATION	PDOC00266

542

DKFZphmcfl\_1a11

-----

group: transmembrane protein

DKFZphmcfl\_1a11 encodes a novel 393 amino acid protein with weak similarity to S.pombe SPBC29A3\_3 protein and S. cerevisiae putative membrane protein YDR255c.

The novel protein contains 1 transmembrane region.  
No informative BLAST results; No predictive prosite, pfam or SCOP motive.

The new protein can find application in studying the expression profile of mammary carcinoma-specific genes and as a new marker for mammary carcinoma cells.

similarity to YDR255c and SPBC29A3.03c

membrane regions: 1

Summary DKFZphmcfl\_1a11 encodes a novel 393 amino acid protein, with similarity to YDR255c and SPBC29A3.03c.

similarity to YDR255c and SPBC29A3.03c

complete cDNA, complete cds, EST hits  
potential start at Bp 110 matches kozak consensus

Sequenced by DKFZ

Locus: /map="542.7 cR from top of Chr5 linkage group"

Insert length: 1819 bp

Poly A stretch at pos. 1808, no polyadenylation signal found

```

1  CCCGGCCCGAG  CCCCCGAAGA  GCCGCCTCAG  CCGGGGGGAG  TTGCTCGGAC
51  TCAAACGTCC  AGTCCTCGTG  CGACCGCGCT  GGGTCGGAAG  TGAGCAGGCT
101  GAGGCCACCA  TGGAGCAGTG  TGGCTGCGTG  GAGAGAGAGC  TGGACAAGGT
151  CCTGCAGAAG  TTCCTGACCT  ACGGGCAGCA  CTGTGAGCGG  AGCCTGGAGG
201  AGCTGTGCA  CTACGTGGGC  CAGCTGCGGG  CTGAGCTGGC  CAGCGCAGCC
251  CTCCAGGGGA  CCCCTCTCTC  AGCCACCCCT  TCTCTGGTGA  TGTACAGTG
301  CTGCCGGAAG  ATCAAAGATA  CGGTGCAGAA  ACTGGCTTCG  GACCATAAGG
351  ACATTACAG  CAGTGTATCC  CGAGTGGGCA  AAGCCATTGA  CAGGAACCTC
401  GACTCTGAGA  TGTGTGGTGT  TGTGTCAGAT  GCGGTGTGGG  ACGCGCGGGA
451  ACAGCAGCAG  CAGATCCTGC  AGATGGCCAT  CGTGGAACAC  CTGTATCAGC
501  AGGGCATGCT  CAGCGTGGCC  GAGGAGCTGT  GCCAGGAATC  AACGCTGAAT
551  GTGGACTTGG  ATTTCAAGCA  GCCTTTCCTA  GAGTTGAATC  GAATCCTGGA
601  AGCCCTGCAC  GAACAAGACC  TGGGTCCTGC  GTTGAATGG  GCCGTCTCCC
651  ACAGGCGAGC  CCTGCTGGAA  CTCAACAGCT  CCCTGGAGTT  CAAGCTGCAC
701  CGACTGCAC  TCATCCGCCT  CTGGCAGGA  GGCCCCGCGA  AGCAGCTGGA
751  GGCCCTCAG  TATGCTCGGC  ACTTCCAGCC  CTTTGCTCGG  CTGCACCAGC
801  GGGAGATCCA  GGTGATGATG  GGCAGCCTGG  TGTACCTCGG  GCTGGGCTTG
851  GAGAAGTCAC  CCTACTGCCA  CCTGCTGGAC  AGCAGCCACT  GGGCAGAGAT
901  CTGTGAGACC  TTTACCCGGG  ACGCCTGTTC  CCTGCTGGGG  CTTTCTGTGG
951  AGTCCCCCT  TAGCGTCAGC  TTTGCCTCTG  GCTGTGTGGC  GCTGCCTGTG
1001  TTGATGAACA  TCAAGGCTGT  GATTGAGCAG  CGGCAGTGCA  CTGGGGTCTG
1051  GAATCACAAG  GACGAGTTAC  CGATTGAGAT  TGAATAGGCC  ATGAAGTGCT
1101  GGTACCACTC  CGTGTTCGCT  TGCCCCATCC  TCCGCCAGCA  GACGTACAGT
1151  TCCAACCCCT  CCATCAAGCT  CATCTGTGGC  CATGTTATCT  CCCGAGATGC
1201  ACTCAATAAG  CTCATTAATG  GAGGAAAGCT  GAAGTGTCCC  TACTGTCCCA
1251  TGGAGCAGAA  CCGGCAGAT  GGGAAACGCA  TCATATTCTG  ATTCTACCT
1301  GGAAGGAATT  TTGTTGAAG  GGGTTTTCAC  CTGTGAGCCT  TGGTCTGTCT
1351  CGGTAGGGTG  GTCAACTTCA  GTGGACTGTG  GTTGGTTTCA  GAGCGCCTGG
1401  CTGAGGAGTT  CCACTGAGGG  GAGCACTGGA  GCAGCCCTTT  GGCAGAGGCT
1451  GAGGAGGGAG  ATGGACCAGC  CCACGCCTGG  CACCTGGCTC  CATGGCATAA
1501  GGAAGGGGAG  ATGCTGGCCT  CTGTGCTCCT  GCTGTCTTTT  CCTGTTTCTG
1551  TTTGCGTTTG  ACTTAGTAGC  AACCGACAGA  GTGGCAAGGG  ATTTGGTCTT
1601  CAGCAGTAGA  CATCCTTCCA  CCCCTGCCCT  CAGCCAAGTC  TCTTGTGCC
1651  ATGCCAATGC  TATGTCACCC  CTGCCCCCTC  GGCCCAAGAG  TGTCCAGCGG
1701  TGGCCACCT  CTTCTCCCA  CTACAGCCTC  AACAGTATGT  ACCATCTCCC
1751  ACTGTAAATA  GTCCCAGTTA  GAACGGAATG  CCGTTGTTTT  ATAACTTTGA
1801  ACAAAATGTA  AAAAAAAAAA

```

## BLAST Results

-----

Entry HS579359 from database EMBL:  
human STS WI-6350.

Score = 1027, P = 9.9e-40, identities = 207/209

## Medline entries

-----

No Medline entry

## Peptide information for frame 2

-----

ORF from 110 bp to 1288 bp; peptide length: 393  
 Category: similarity to unknown protein

```

1 MEQCACVERE LDKVLQKFLT YGQCERSLE ELLHYVGQLR AELASAAQOG
51 TPLSATLSLV MSQCCRKIKD TVQKLASDHK DIHSSVSRVG KAIDRNFDE
101 ICGVVSDAVW DAREQQQIL QMAIVEHLYQ QGMLSAEEL CQESTLNVDL
151 DFKQPFLELN RILEALHEQD LGPALEWAVS HRQRLLELNS SLEFKLHRLH
201 FIRLLAGGPA KQLEALSYAR HFQPFARLHQ REIQVMGSL VYLRGLGLEKS
251 PYCHLLDSSH WAEICETFTF DACSLGLSV ESPLSVSFAS GCVALPVLMM
301 IKAVIEQRQC TGVWNHKLDEL PIEIELGMKC WYHSVFACPI LRQQTSDSNP
351 PIKLICGHVI SRDALNKLIN GGLKPCPYCP MEQNPADGKR IIF

```

## BLASTP hits

No BLASTP hits available

## Alert BLASTP hits for DKFZphmcf1\_lal1, frame 2

TREMBL:SPBC29A3\_3 gene: "SPBC29A3.03c"; product: "hypothetical protein"; S.pombe chromosome II cosmid c29A3., N = 2, Score = 302, P = 3.4e-42

PIR:S67312 probable membrane protein YDR255c - yeast (Saccharomyces cerevisiae), N = 1, Score = 271, P = 5.3e-22

TREMBL:CET07D1\_2 gene: "T07D1.2"; Caenorhabditis elegans cosmid T07D1., N = 1, Score = 193, P = 5.6e-13

>TREMBL:SPBC29A3\_3 gene: "SPBC29A3.03c"; product: "hypothetical protein"; S.pombe chromosome II cosmid c29A3.  
 Length = 398

## HSPs:

Score = 302 (45.3 bits), Expect = 3.4e-42, Sum P(2) = 3.4e-42  
 Identities = 55/142 (38%), Positives = 89/142 (62%)

Query: 252 YCHLLDSSHWAIEICETFTFTRDACSLLGLSVESPLSVSFASGCVALPVLMMIKAVIEQRQCT 311  
 Y +LD W + F R+ C+ LG+S+ESPL + +G +ALP+L+ + +++++  
 Sbjct: 258 YIDVLDLD-WKSLELFLVREFCAALGMSLESPLDIVVNAGAIAPILLKMSSIMKKKHT 316

Query: 312 GVVNNHKLDELPIEIELGMKCWYHSVFACPI LRQQTSDSNPPIKLICGHVISRDALNKLING 371  
 W + ELP+EI L +HSVF CP+ +Q ++ NPP+ + CGHVI +++L +L  
 Sbjct: 317 --WTSQGELPVEIFLPSSYHFSVFTCPVSKEQATEENPPMMSCGHVIVKESLRQLSRN 374

Query: 372 G--KLKPCPYCPMEQNPADGKR IIF 393  
 G + KCPYCP E AD R+ F  
 Sbjct: 375 GSQRFKPCPYCPNENVAADAIRVYF 398

Score = 161 (24.2 bits), Expect = 3.4e-42, Sum P(2) = 3.4e-42  
 Identities = 51/221 (23%), Positives = 102/221 (46%)

Query: 22 GQCERSLEELLHYVGQLRAELASAAQOGTPLSATLSLVMSQCCRKIKD TVQKLASDHK 81  
 G C L EL + + + L+ P ++ LV C K + L K  
 Sbjct: 15 GNKCLAKLNEL---ESILKDAKSKLKD-PTSMKELVA--CSEKQQVFDLKRTEKK 67

Query: 82 IHSSVSRVGKAIDRNFDEICGVVSDAVWDAREQQQILQMAIVEHLYQQGMLSAEELC 141  
 H+S++R GK +++ F+ ++ + +++++++ + A+ H ++QG + +A C  
 Sbjct: 68 FHTSLNRFGKTLEKKFNFDLEDIKLHSSFESKRE---IDTALS LHF FRQGDVELAHLFC 124

Query: 142 QESTLNVDLDFKQPFLELNRIEALHEQDLGPALEWAVSHRQRLLELNS SLEFKLHRLHF 201  
 +E+ + + F L I++ ++DL +EWA R L SSLE+ L +  
 Sbjct: 125 KEAGIEEPSSELHVFTLLKSI VQGI RDKDLKPIEWASQCRGYLERKGS SLEYTLQKYRL 184

Query: 202 IRLLAGGPAKQL-EALSYAR-HFQPFARLHQREIQVMGSLVY 242  
 + K + A+ Y R + F + H +IQ M +L +



Pedant information for DKFZphmcfl\_1a11, frame 2

## Report for DKFZphmcf1 1a11.2

[illegible]

Prosites for DKF2phmcf1\_1a11.2

PS00001	189->193	ASN_GLYCOSYLATION	PDOC00001
PS00005	180->183	PKC_PHOSPHO_SITE	PDOC00005
PS00006	28->32	CK2_PHOSPHO_SITE	PDOC00006
PS00006	135->139	CK2_PHOSPHO_SITE	PDOC00006
PS00006	190->194	CK2_PHOSPHO_SITE	PDOC00006
PS00007	211->219	TYR_PHOSPHO_SITE	PDOC00007
PS00007	27->36	TYR_PHOSPHO_SITE	PDOC00007
PS00007	244->253	TYR_PHOSPHO_SITE	PDOC00007
PS00008	37->43	MYRISTYL	PDOC00008
PS00008	50->56	MYRISTYL	PDOC00008
PS00009	387->391	AMIDATION	PDOC00009
PS00013	282->293	PROKAR_LIPOPROTEIN	PDOC00013

545

DKFZphmcf1\_1c23

-----

group: mammary carcinoma derived

DKFZphmcf1\_1c23.1 encodes a novel 311 amino acid proline rich protein.

No informative BLAST results; No predictive prosite, pfam or SCOP motive.

The new protein can find application in studying the expression profile of mamma carcinoma-specific genes.

unknown, proline rich protein

complete cDNA, complete cds? potential start at Bp 50, EST hits

Sequenced by DKFZ

Locus: unknown

Insert length: 3077 bp

Poly A stretch at pos. 3067, polyadenylation signal at pos. 3048

```

1 AACTGGCCCC CTCCCCACC CCCTGCCCCT GAGGAGCAGG ACCTGTCCAT
51 GGCTGACTTC CCCCCACCAG AGGAGGCTTT TTTCTCTGTG GCCAGCCCTG
101 AGCCTGCAGG CCTTCAGGCG TCCCAGAGCG TTGTGAGCTC CCGGCTGCT
151 TCCTCTCTCT CAGCTACTGC TTTGAGATT CAGCCCCCGG GTAGCCAGAG
201 CCCTCTCTCA GCTCCGCCAG CCCCAGCTCC TGCTAGTTCC GCCCCAGGGC
251 ATGTGGCCAA GCTCCCTCAG AAGGAACCGG TGGGCTGTAG CAAGGGTGGT
301 GGGCCTCCCA GGGAGGACGT AGGTGCGCCC CTGGTCACGC CCTCGCTCCT
351 GCAGATGGTG CGGCTGCGCT CCGTGGGTGC TCCAGGAGGG GCTCCACCCC
401 CAGCACTGGG GCCATCGGCC CCCCAGAAAC CACTGCGAAG GGCCCTGTCA
451 GGGCGGGCCA GCCAGTGCC TGCCCCCTCC TCAGGGCTCC ATGCTGCGGT
501 CCGACTCAAG GCCTGCAGCC TGGCCGCCAG TGAAGGCTC TCAAGTGCTC
551 AGCCCAACGG ACCGCTGAG GCAGAGCCAC GGCTCCCCA GTCCCTGCC
601 TCAACGGCCA GTTTCATCTT CTCCAAGGGC TCTAGGAAGC TGCAGCTGGA
651 GCGGCCCGTG TCCCCTGAGA CCCAGGCTGA CCTCCAGCGG AATCTGGTGG
701 CAGAACTCCG GAGCATCTCA GAGCAGCGGC CACCCAGGCC CCCAAAGAA
751 TCACCTAAGG CTCCCCCACC TGTGGCCCGC AAGCCGTCTG TGGGAGTCCC
801 CCCACCCGCC TCCCCCAGTT ACCCTCGAGC TGAGCCCTTT ACTGCTCCTC
851 CCACCAATGG GCTCCCTCAC ACCCAGGACA GGAATAAGAG GGAGCTGGCG
901 GAGAATGGAG GTGTCTGCA GCTGGTGGGC CCAGAGGAGA AGATGGGCCT
951 CCGGGCTCA GACTCACAGA AAGAGCTGGC CTGACCACCA GGCACCTCAC
1001 TGGCACTGCT GACCATCCC AGAAACAAA TCTCAGGGAC CCGAGCAGT
1051 CCAAGGACGA GAGGATACAG CAGACACAA CTAATAGAGA GGGCGCCTGC
1101 AGCCTTAACC TCCACGGCCT TCGATACTTA TGCAAGCCTG GTGTTGCTCC
1151 TGTCTCTAGA GTCATCTGCG GCTCATGCTT TTTCCCGAAT GGGTTACCT
1201 CTGGCAGTTG CCGCTTCAGT CTTGGCCTTA GCCTCATCTT GAAGTGGGTA
1251 GCTGGCGGGA GAGGGTGGCT GCGCCCCCTG CTGGCCCTGA GGCTGCAGAG
1301 TTGGGAGCAG GACACCTCAC CTGAGTTTCA TTTTTCATCA TGTCCAAACC
1351 ATGCACATAC TATAGTCCAG AATCAAAGCA CTTTGAATAA GTGGCTGCAT
1401 GGCATCTCTC CAGGGCCAGG GAAGTTGATC TCCAAGGGCC TGTTTACATG
1451 GCAGCAGAAAT CCATCCCCGG CAGTCAGCCC ATAGCTTGGG ACCAGTCTGT
1501 GCCTCTCTCG CCAGTCCAGT TTACTCTCTT TGGTCTCTGA AGGTGGCCAA
1551 GTCATTGTGT TCCCACAGGC TTCTTAGGC TGGGGCAGG TGTGGGCTG
1601 TGGAAATCCA AAGCACAAA GGTGCAGAGG GGATTGGCCT TCCTGTGCCT
1651 CAACCTACCA ACCACCTCC TGCCCTCCAG TTCTGCCAGG TGCTCCATGC
1701 TGGGGACAAG TAGGAGACTG CCAGGGCCCA AAGAAATGGG TGAGCAGTAG
1751 AGTCATCTCG GGGCACTTGG CAGTGTCAAG CACCTGCCCC TTGCTCTCTT
1801 GACCACACTG GGGTGGGTGG GCCCCAGCA CTTACAGAGG AGGAGCCTTT
1851 GGGGTGAGCA AGCACTGAGG AGGTGGATGG AAGGGAGCAT CTGGAGGGGG
1901 GGAGCTTCTT TGAGCAGTGG GCCCAGGCCT GGCCCTCCAC ACTTCATTCT
1951 CTGACCTTTC TCTCTCCTCA TTTGGGTGCA TGTCTTTCT GCAGCTGCCT
2001 TTCAGCACAG GTGTTCCAC TGGGGCAGC TAACGCTGAG TGACAAGGAT
2051 GGAAGCCAC AGGTGCATTT TACTCAAGTC TTCTCTAGTC AATGAGGGGC
2101 ACCCAGTGCT TTAGGGCAG GCTGGGTGGT GGTCCCTTAG GTATCAGCCT
2151 CTCTTACTGT ACTTCCGGG AATGTTAACC TTTCTATTTT CAGCCTGTGC
2201 CACCTGTCTA GGAAGCTGG CTTCCTCATT GGCCCTGTG GGTCCACAGC
2251 AGCGTGGCTG CCCCCAGGG CCACCGCTTC TTTCTTGATC CTCTTTCCTT
2301 AACAGTGACT TGGGCTTGAG TCTGGCAAG AACCTTGCTT TTAGCTTCAC
2351 CACCAAGGAG AGAGGTGAC ATGACCTCCC CGCCCCCTCA CCAAGGCTGG
2401 GAACAGAGGG GATGTGGTGA GAGCCAGGTT CCTCTGGCCC TCTCCAGGGT
2451 GTTTTCCACT AGTCACTACT GTCTTCTCT TGTAGTAAT CAATCAATAT
2501 TCTTCCCTTG CTTGTGGGCA GTGGAGAGTG CTGCTGGGTG TACGCTGCAC
2551 CTGCCCCTAG AGTTGGGGAA AGAGGATAAT CAGTGAGCAC TGTTCCTGCT
2601 AGAGCTCTTG ATCTACCCA CCCCCTAGGA TCCAGGACTG GGTCAAAGCT
2651 GCATGAAACC AGGCCCTGGC AGCAACCTGG GAATGGCTGG AGGTGGGAGA
2701 GAACCTGACT TCTTTTCCC TCTCCTCTCT CCAACATTAC TGGAACTCTA

```

```
2751 TCCTGTTAGG ATCTTCTGAG CTTGTTTCCC TGCTGGGTGG GACAGAGGAC
2801 AAAGGAGAAG GGAGGGTCTA GAAGAGGCAG CCCTTCTTTG TCCTCTGGGG
2851 TAAATGAGCT TGACCTAGAG TAAATGGAGA GACCAAAAGC CTCTGATTTT
2901 TAATTTCCAT AAAATGTTAG AAGTATATAT ATACATATAT ATATTTCTTT
2951 AAATTTTGA GTCTTTGATA TGCTTAAAAA TCCATTCCCT CTGCCCTGAA
3001 GCCTGAGTGA GACACATGAA GAAACTGTG TTTTATTAA AGATGTTAAT
3051 TAAATGATTG AAACCTGAAA AAAAAAA
```

## BLAST Results

No BLAST result

## Medline entries

No Medline entry

## Peptide information for frame 1

ORF from 49 bp to 981 bp: peptide length: 311  
Category: putative protein  
Classification: unset

```
1 MADFPPEEA FFSVASPEPA GPSGSPVLVS SPAASSSSAT ALQIQPPGSP
51 DPPPAPPAPA PASSAPGHVA KLPQKEPVGC SKGGGPPRED VGAPLVTPSL
101 LQMVRLRSVG APGGAPTAL GPSAPQKPLR RALSGRASP PAPSSGLHAA
151 VRLKACSLAA SEGLSSAQPN GPPEAEPRPP QSPASTASFI FSKGSRKLQL
201 ERVSPETQA DLQRNLVAEL RSISEQRPPQ APKKSPKAPP PVARKPSVGV
251 PPPASPSYPR AEPLTAPPTN GLPHTQDRTK RELAENGVL QLVGPPEKMG
301 LPGSDSQKEL A
```

## BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphmcf1\_1c23, frame 1

PIR:S49915 extensin-like protein - maize, N = 1, Score = 215, P = 6.1e-15

PIR:A28996 proline-rich protein M14 precursor - mouse, N = 1, Score = 191, P = 3.8e-13

>PIR:S49915 extensin-like protein - maize  
Length = 1,188

## HSPs:

Score = 215 (32.3 bits), Expect = 6.1e-15, P = 6.1e-15  
Identities = 81/269 (30%), Positives = 115/269 (42%)

```
Query: 5 PPPEEAFFS---VASPEPAGPSGSPVLVSSPAASSSSATALQIQPPGSP--DPPP---A 55
      PPP S V SP P P SP PA +SS ++ PP +P PPP +
Sbjct: 598 PPPPAPVASPPPPVKSPPPTPVASPP---PPAPVASSPPPMKSPPTPVSSPPPPPEKS 654

Query: 56 PPAPAPASSAPGHVAKLPQKEPVGC SKGGGPPREDVGAPLVTPSLLQMVRLRSVGAPGGA 115
      PP P PA S P + P P K PP + + P + PS + P
Sbjct: 655 PPPPPAKSTPPP-EEYPT--PPTSVKSSPPPEKSLPPPTLIPSPPPQEKPTPPSTPSKP 711

Query: 116 PTPALGPSAPQKPLRRA-LSGRASVPVAPSSGLHAAVRLKACSLAASEGLSSAQPNGPPE 174
      P+ PS P++P+ + ++SP PAP S +LA S + + PP
Sbjct: 712 PSSPEKPSPPKEPVSSPPQTPKSSPPAPVSSPPPTPVSSPPALAPVSSPPSVKSSPPPA 771

Query: 175 AEPRPPQSPASTASFIFSKGSRKLQLERPV-SPETQADLQRNLVAELRSISEQRPPQAPK 233
      PP +P +S +Q+ P +P++ L V+ + + PP AP
Sbjct: 772 PLSSPPAPQVKSS-----PPPVQVSSPPAPKSSPPLAP--VSSPPQVEKTSPPPAFL 823

Query: 234 KSPKAPPPVARKPSVGV--PPPASPSYPRAEPLTAPPTNGLP 273
      SP P + P V V PPP S P P+++PP P
Sbjct: 824 SSPPLAPK-SSPHVVVSSPPVVKSSPPAPVSSPPLTPKP 864

Score = 206 (30.9 bits), Expect = 9.1e-14, P = 9.1e-14
```

Identities = 82/261 (31%), Positives = 108/261 (41%)

Query: 17 PEPAG-PSGSPVLSSPAASS---SSATALQIQPPGSPDPPAP---PAPAPASSAPGHV 69  
 P P G P S P + PAAS+ S T + P P+P P P P P +P  
 Sbjct: 410 PTPGGGPPSSP-VPGKPAASAPMPSHTPPDVSPLEPEPSVPAPAPMPMPTPHSPPAD 468

Query: 70 AKLPQKEPV-GCSKGGGPPREDVGAPLVTPLSLQMVRLRSVGAPGGAPTALGSPAPQKP 128  
 +P PV G S P V P + +V+L AP G+P P + ++P P  
 Sbjct: 469 DYVPPTPPVPGKSPATSPSPQVQPPAASPPPSLVKLSPPQAPVGSPPPPVKTTSPPAP 528

Query: 129 LRRALSGRASVPAPSSGLHAAVRLKACSLAASEGLSSAQPNGPPEAEPRPPQSPASTAS 188  
 + G S P P S + +K+ A G + P PPE P PP AS  
 Sbjct: 529 I-----GSPSP-PPPVSVSPPPPVKSPPPAPVG---SPP--PPEKSPPPAPVASPP 577

Query: 189 FIFSKGRKLQLERPVSPETQADLQRLNVAELRSISEQRPPQAPKKSPPKAPPPVARKPS- 247  
 + S L P P ++ VA + PP P SP P PVA P  
 Sbjct: 578 PVKSPPPPTLVASPP--PPVKSPPPPAPVASPPPVKSPPPPTPVASPPPPAPVASSPP 635

Query: 248 VGVPPP---ASPSYPRAEPLTAPPTNGLPHTQD 277  
 + PPP +SP P P PP P ++  
 Sbjct: 636 MKSPPPTPVSSPPPEKSPPPPPAKSTPPPEE 669

Score = 202 (30.3 bits), Expect = 2.9e-13, P = 2.9e-13  
 Identities = 81/254 (31%), Positives = 110/254 (43%)

Query: 16 SPEPAGPSGSPVLSSPAASS---SSATALQIQPPGSPDPPAPAPAPASSAPGHVA 70  
 SP PA P SP L SSP SS ++ PP +P PP P PA S P HV+  
 Sbjct: 817 SPPPA-PLSSPPLAPKSSPPHVVSPPPPVKSPPPPAPVSSPPLTPKPA---SPPAHVS 872

Query: 71 KLPQ---KEPVGCSKGGGPPREDVGAPLVTPLSLQMVRLRSVGAPGGAPTALGSPAPQ 126  
 P+ P + PP E +P TP L ++S P +P + P +  
 Sbjct: 873 SPPEVVKPSTPPAPTTVISPPSEPKSSPPPTPVSLPPPIVKSSPPAMVSSPMTPKSSP 932

Query: 127 KPLRRAL---SGRASVPAPSSGLHAAVRLKACSLAASEGLSSAQPNGPPEAEPRPPQSP 183  
 P+ + + ++SP PAP S A K+ A L P PPE + PP +P  
 Sbjct: 933 PPVVSSPPPTVKSSPPAPVSSPPATP--KSSPPAPVNL---P--PPEVKSPPPTP 984

Query: 184 ASTASFIFSKGRKLQLERPVSPETQADLQRLNVAELRSISEQRPPQAPKKSPPKAPPPVA 243  
 S+ + P PE ++ V+ + PP AP SP PPPV  
 Sbjct: 985 VSSPPAPKSSPPAPMSSPPPEVKSPPPPAPVSSPPPVKSPPPAPVSSP--PPPVK 1042

Query: 244 RKPS---VGVPPPASPSPYRAEPLTAPP 268  
 P V PPP S P P++PP  
 Sbjct: 1043 SPPPPAPVSSPPPVKSPPPAPISPP 1070

Score = 190 (28.5 bits), Expect = 7.9e-12, P = 7.9e-12  
 Identities = 74/264 (28%), Positives = 111/264 (42%)

Query: 5 PPPEEAFFSVASPEPAGPSGSPVLSSPAAS---SSATALQIQPPGSPDPPAPAPAPAS 63  
 PPP S PE + P P + P + T+++ PP PP P+P  
 Sbjct: 639 PPPPTVSSPPPEKSPPPPPAKSTPPPEEYPTPTSVKSSPPPEKSLPPTLIIPSP 698

Query: 64 SAPGHVAKLPQKEPVGCSKGGGPPREDVGAPLVTPLSLQMVRLRSVGAPGGAPTALGSP 123  
 P K P K PP+E V +P TP V +P PTP P  
 Sbjct: 699 QEKPTPPTSPKPPSSPEKPS-PPKEPVSSPPQTPK--SSPPAPVSSP--PPTPVSSP 753

Query: 124 APQKPLRRALSGRASVPAPSSGLHAAVRLKACSLAASEGLSSAQPNGPPEAEPRPPQSP 183  
 A P+ S ++SP PAP S A ++K+ + + P PP + PP +P  
 Sbjct: 754 A-LAPVSSPPSVKSSPPAPLSSPPAPQVKS----SPPPVQVSSP--PPAPKSSPPLAP 806

Query: 184 ASTASFIFSKGRKLQLERP-VSPETQADLQRLNVAELRSISEQRPPQAPKKSPPKAPPPV 242  
 S+ + L P ++P++ +V+ + + PP AP SP P  
 Sbjct: 807 VSSPPQVEKTSPPAPLSSPPLAPKSSPP--HVVVSSPPPVKSSPPAPVSSPPLTPK 864

Query: 243 ARKPS-VGVPPP---PASPSYPR-----AEPLTAPP 268  
 A P+ V PP P++P P +EP ++PP  
 Sbjct: 865 ASPPAHVSSPPEVVKPSTPPAPTTVISPPSEKSSPP 901

Score = 189 (28.4 bits), Expect = 1.0e-11, P = 1.0e-11  
 Identities = 86/271 (31%), Positives = 112/271 (41%)

Query: 5 PPPEEAFFSVASPEPAGPSGSPVLSSPAAS---SSATALQIQPPG---SPDPPAP--- 56  
 PPP A S P P S P + VSSP A SS A PP PPPAP  
 Sbjct: 768 PPP--APLSSPPAPQVKSSPPPVQVSSPPAPKSSPPLAPVSSPPQVEKTSPPAPLSS 825

Query: 57 PAPAPASSAPGHVAKLPQKEPVGCSKGGGPPREDVGAPLVTPLSLQMVRLRSVGAPGGAP 116  
 P AP SS P V P PV S PP V +P +TP V +P  
 Sbjct: 826 PPLAPKSSPPHVVSPP--PVVKSS---PPAPVSSPPLTPKPPASPPA--HVSPPPEVV 878

Query: 117 TPALGPSAPQKPLRRALSGRASVPAPSSGLHAAVRLKAC-SLAASEGL---SSAQ--- 169  
 P+ P AP + ++SP P P S V+ ++ +S + SS P

Sbjct: 879 KPST-PPAPTTVISPPSEKSSPPPTPVSLPPPVIKSSPPAMVSSPMTPKSSPPV 937

Query: 170 -NGPPEAEPRPPQSPASTASFIIFSKGSRKLQLERPVSPETQADLQRLNVAELRSISEQRP 228  
+ PP + PP +P S+ + P PE ++ V+ + P

Sbjct: 938 SSSPPTVKSSPPAPVSSPPATPKSSPPAPVNL P-PPEVKSSPPPTPVSSPPAPKSSP 996

Query: 229 PQAPKKSAPKPPPVARKPS---VGVPASPSPYPRAEPLTAPP 268  
P AP SP PPP + P V PPP S P P+++PP

Sbjct: 997 PPAPMSSP--PPPEVKSSPPAPVSSPPPVKSSPPAPVSSPP 1038

Score = 181 (27.2 bits), Expect = 8.8e-11, P = 8.8e-11  
Identities = 73/277 (26%), Positives = 105/277 (37%)

Query: 3 DFPPEEAEFFSVASPEPAGSGSPSELVSSPAASSSSATALQIQPP---GSPDPP---PA 55  
D+ PP V P S SP+ V PAAS+ + +++ PP GSP PP +

Sbjct: 469 DYVPPTTP---VPGKSPATSPSPQ-VQPPAASTPPPSLVKLSPPQAPVGSPPPVKTS 524

Query: 56 PPAPASSAPGHVAKL---PQKEPVGCSKGGGPPREDVGAPLVTPLQMVRLRSVGA 111  
PPAP + S P V+ + P K P + G PP + P P ++S

Sbjct: 525 PPAPIGSPSPPPVSVVSSPPPVKSSPPAPVGSPPPEKSPPPAPVASPPPVKSSPP 584

Query: 112 PG--GAPTPALGPSAPQKPLRRA---LSGRASVPAPSSGLHAAVRLKACSLAASEGLSS 166  
P +P P + P P+ + P P S A V + + +

Sbjct: 585 PTLVASPPPVKSSPPAPVASPPPVKSSPPPTPVASPPAPVASSPPPMKSPPPPTP 644

Query: 167 AQPNGPPEAEPRPPQSPASTASFIIFSKGSRKLQLERPVSPETQADLQRLNVAELRSISEQ 226  
PPE P PP PA + + ++ PE L+ +

Sbjct: 645 VSSPPPEKSP-PPPPAKSTPPPEEYPTPTTSVKSSPPPEKSLP-PPTLIPSPPPQEK 702

Query: 227 RPPQAPKKSAPKPP-PVARKPSVGVPPASPSPYPRAEPLTAPP 268  
PP P K P +P P K V PP S P P+++PP

Sbjct: 703 TTPSTPSKPPSSPEKSPPEKSPVSSPPQTPKSSPPAPVSSPP 745

Score = 177 (26.6 bits), Expect = 2.6e-10, P = 2.6e-10  
Identities = 78/264 (29%), Positives = 105/264 (39%)

Query: 5 PPPEEAEFFSVASPEPAGP---SGSPSELVSSPAASSSSATALQIQPPGSP--DPPAP-- 56  
PPP +P+PA P S PE+V P+ + T I PP P PPP P

Sbjct: 850 PPAPVSSPPLTPKPASPPAHVSSPPEVVK-PSTPPAPTTP--ISPPSEKSSPPPTPV 906

Query: 57 -PAPAPASSAPGHVAKLPQKEPVGCSKGGGPPREDVGAPLVTPLQMVRLRSVGA 115  
P P SS P + P P P PP V +P P++ V +P

Sbjct: 907 LPPPIVKSSPPAMVSSPMTPKS-----SPPVVVSSP--PPTVKSSPPAPVSSPPAT 959

Query: 116 PTPALGPSAPQKPLRRALSGRASVPAPSSGLHAAVRLKACSLAASEGLSSAQNGPPEA 175  
P + P+ P ++SP P P S A + S +SS P PPE

Sbjct: 960 KSSPPAPVNLPPPEV---KSSPPPTPVSSPPAPK----SSPPAPMSSP-P--PPEV 1009

Query: 176 EPRPPQSPASTASFIIFSKGSRKLQLERPVSPETQADLQRLNVAELRSISEQRPQAPKKS 235  
+ PP +P S+ + P P ++ V+ + PP AP S

Sbjct: 1010 KSPPPAPVSSPPPVKSSPPAPVSSP-PPPVKSSPPAPVSSPPPVKSSPPAPISS 1068

Query: 236 PKAPPPVARKPS---VGVPASPSPYPRAEPLTAPP 268  
P PPPV P V PPP S P P+++PP

Sbjct: 1069 P--PPPVKSSPPAPVSSPPPVKSSPPAPVSSPP 1102

Score = 177 (26.6 bits), Expect = 2.6e-10, P = 2.6e-10  
Identities = 82/267 (30%), Positives = 110/267 (41%)

Query: 17 PEPAG-PSGSELVSSPAASS---SSATALQIQPPGSPDPPAP---PAPAPASSAPGHV 69  
P P G P SP + PAAS+ S T + P P+P P P P P +P

Sbjct: 410 PTPGGGPPSSP-VPGKPAASAPMPSHTPPDVSPELPEPSVPVAPAPMMPMTPHSPPAD 468

Query: 70 AKLPQKEPV-GCSKGGGPPREDVGAPLVTPLQMVRLRSVGA 128  
+P PV G S P V P + +V+L AP G+P P + ++P P

Sbjct: 469 DYVPPTTPVPGKSPATSPSPQVQPPAASTPPPSLVKLSPPQAPVGSPPPVKTSPPAP 528

Query: 129 LRRALSGRASVPAPSSGLHAAVRLKACSLAASEGLSSAQNGPPEAEPRPPQSPASTAS 188  
+ G SP P P S + +K+ A G + P PPE P PP AS

Sbjct: 529 I-----GSPSP-PPPVSVVSSPPPVKSSPPAPVG---SPP--PPEKSSPPAPVASPP 577

Query: 189 FIFSKGSRKLQLERPV---SPETQADLQRLNVAELRS-----ISEQRPQA-----PK 233  
+ S L P SP A + + ++S ++ PP P

Sbjct: 578 PVKSSPPPTLVASPPPVKSSPPAPVA-SPPPVKSSPPPTPVASPPAPVASSPPPM 636

Query: 234 KSPKAPPPVARKP---SVGVPPASPSPYPRAEPLTAPPTN 270  
KSP P PV+ P PPP + S P E PPT+

Sbjct: 637 KSPPTPVSSPPPEKSSPPPPPAKSTPPPEEYPTPTTS 676

Score = 170 (25.5 bits), Expect = 1.6e-09, P = 1.6e-09  
Identities = 78/279 (27%), Positives = 108/279 (38%)

Query: 5 PPPEEAFSVASPEPAGPSGSPVLVSSPAASSSSATALQIQQSPDPPAPAPASS 64  
 PP S S + P + P + P SS A+ PP +P +PP P SS  
 Sbjct: 883 PPAPTTVISPPSEPKSSPPPTPVSLPPPIVKSSPPAPMVSSPMTPKS--SPP-PVVVSS 939

Query: 65 APGHVAKLPQKEPVGCSKGGPPREDVGAPLVTPLSLQMVRLRSVGAPG--GAPTALGP 122  
 P V P PV PP +P P L ++S P +P PA  
 Sbjct: 940 PPFTVKSSPPAPVS-----SPPATFKSSPPAPVNLPPPEVKSSPPPTPVSSPPAPKS 994

Query: 123 SAPQKPLRRALSG--RASVPVAPSSGLHAAVRLKACSLAASEGLSSAQNGPPEAEPRPP 180  
 S P P+ ++ P PAP S V+ S +SS P PP + PP  
 Sbjct: 995 SPPAPMSSPPPEVKSSPPAPVSSPPPVK----SPPAPVSS--P--PPPVKSPP 1046

Query: 181 QSPASTASFIKSGSRKLQLERPVSPETQADLQRLVAELRSISEQRPPQAPKSPKAPP 240  
 +P S+ + P P ++ V+ + PP AP SP PP  
 Sbjct: 1047 PAPVSSPPPVKSPPPPAPISSP-PPPVKSPPAPVSSPPPVKSPPPPAPVSSP--PP 1103

Query: 241 PVARKPS---VGVPPPAS---PSYPRAEPLTAPPTNGLPHTQDRTKREL 283  
 P+ P V PPPA TS P P+++PP P + ++ L  
 Sbjct: 1104 PFKSPPPPAPVSSPPAPVKPSLPPAPVSSPPPVTPAPPKKEEQL 1152

Score = 169 (25.4 bits), Expect = 2.1e-09, P = 2.1e-09  
 Identities = 75/266 (28%), Positives = 104/266 (39%)

Query: 3 DFPPEEAFSVASPEPAGPSGSPVLVSSPAASSSSATALQIQQP----GSPDPP---PA 55  
 D+ PP V P S SP+ V PAAS+ + +++ PP GSP PP +  
 Sbjct: 469 DVPPTTP---VPGKSPFATSPSPQ-VQPPAASTPPSLVKLSPPQAPVGSPPPVKTTS 524

Query: 56 PPAPAPASSAPGHVAKLPQKEPVGCSKGGPPREDVGAPLVTPLSLQMVRLRSVGAPGGA 115  
 PPAP + S P V+ + PV PP VG+P P V+P  
 Sbjct: 525 PPAPIGSPSPPPVSVVSSPPPVKSP---PPAPVGSPP---PPPEKSPPPAPVASP--- 575

Query: 116 PTPALGPSAPQKPLRRALSGRASVPVAPSSGLHAAVRLKACSLAASEGLSSAQNGPPEA 175  
 P P P P ++ P PAP + V+ S ++S P P +  
 Sbjct: 576 PPPVKSPPPTLVASPPPVKSPPPPAPVASPPPVK----SPPPTPVASPPPPAPVAS 631

Query: 176 EPRPPQSPASTASFIKSGSRKLQLERPVSPETQADLQRLVAELRSISEQRPPQAPKKS 235  
 P P +SP K P P S+ PP+  
 Sbjct: 632 SPPPMKSPPPPTPVSSPPPEKSP---PPPPAKSTPPPEYTPPTSVKSSPPPEKSLPP 689

Query: 236 PK---APPPVARK--PSVGVPPPASPSYPRA--EPLTAPP 268  
 P +PPP + PS PP+SP P EP+++PP  
 Sbjct: 690 PTLIPSPPPQEKPTPSTPSKPFSSPEKPSPPKEPVSSPP 729

Score = 168 (25.2 bits), Expect = 2.7e-09, P = 2.7e-09  
 Identities = 75/267 (28%), Positives = 102/267 (38%)

Query: 2 ADFPPEEAFSVASPE-PAGPSGSPVLVSSPAASSSSATALQIQQSPDPP-PAPPAP 59  
 A PPP + ++ P+ P G P +SP A S + SP PP +PP P  
 Sbjct: 496 ASTPP---SLVKLSPPQAPVGSPPPVKTTSPPAPIGSPSPPPVSVSSPPPVKSPPPP 553

Query: 60 APASSAPGHVAKLPQKEPVGCSKGGPPREDVGAPLVTPLSLQMVRLRSVGAPGAPTPA 119  
 AP S P P PV PP + P + S V+ AP +P P  
 Sbjct: 554 APVGSPPPEKSPPPAPVASPP---PPVKSPPPTLVASPPPVKSPPPPAPVASPPPP 610

Query: 120 LGFSAPQKPLRRALSGRASVPVAPSSGLHAAVRLKACSL-AAASEGLSSAQNGPPEAEPR 178  
 + P P+ + P PAP + ++ +S P PP A+  
 Sbjct: 611 VKSPPPPTVA-----SPPPPAPVASSPPPMKSPPPPTPVSSPPPEKSPPPPPPAKST 664

Query: 179 PP--QSPASTASFIKSGSRKLQLERP---SPETQADLQRLVAELRSISEQRPPQAPK 233  
 PP + P S S K L P SP Q S ++P +P  
 Sbjct: 665 PPPEEYTPPTSVKSSPPPEK-SLPPPTLIPSPPPQEKPTPSTPSKPFSSPEK--SPP 721

Query: 234 KSPKAPPPVARKPSVGVPPPASPSYPRAEPLTAPP 268  
 K P + PP K S PPPA S P P+++PP  
 Sbjct: 722 KEPVSSPPQTKSS---PPAPVSSPPPTPVSSPP 753

Score = 166 (24.9 bits), Expect = 4.6e-09, P = 4.6e-09  
 Identities = 81/268 (30%), Positives = 108/268 (40%)

Query: 5 PPPEEAF---FSVASPEPAGPSGSP-LLVSSPAASSSS---ATALQIQQSPDPP-- 54  
 PPPE++ VASP P S P LV+SP S A PP PPP  
 Sbjct: 560 PPPEKSPPPAPVASPPPVKSPPPTLVASPPPVKSPPPPAPVASPPPVKSPPPPPTP 619

Query: 55 --APPAPAPASSAPGHVAKLPQKEPVG---SKGGPPREDVGAPLVTPLSLQMVRLRS 108  
 +PP PAP +S+P + P PV K PP P ++S  
 Sbjct: 620 VASPPPPAPVASSPPPMKSPPPPTPVSSPPPEKSPPPPPPAKSTPPPEEYTPPTSVKS 679

Query: 109 VGAPGGA-PTPALGPSAPQKPLRRALSGRASVPVAPSSGLHAAVRLKACSLAASEGLSSA 167  
 P + P P L PS P P + + ++P PSS + + S SS  
 Sbjct: 680 SPPPEKSLPPPTLIPSP--PQEKPTPSTPSKPFSSPEKPSPPKEPVSSPPQTKSSP 736

Query: 168 QPNGPPEAEPRPPQSPASTASFIFSKGSRKLQLERPVSPETQADLQRLNVAELRSISEQR 227  
 P P P SP + A + S S K P + P + + + +  
 Sbjct: 737 PPAPVSSPPPTPVSSPPALAP-VSSPPSVKSS--PPAPLSSPPAPQVKSSPPPVQVSS 793

Query: 228 PPQAPKKSPPKAPPPVARKPSVGVPPPASPSYPRAEPLTAPP 268  
 PP APK SP P+A P V PP + P PL++PP  
 Sbjct: 794 PPPAPKSSP----PLA--P-VSSPPQVEKTSPPAPLSSPP 827

Score = 165 (24.8 bits), Expect = 6.0e-09, P = 6.0e-09  
 Identities = 79/264 (29%), Positives = 105/264 (39%)

Query: 5 PPPEEAFSVASPEPAG-PSGSP--ELVSSPAASSSSATALQIQQPGSPDPPPP-APPAPA 60  
 PPP + + + P P G PS P +VS P S P GSP PP +PP PA  
 Sbjct: 517 PPPVK---TTSPPAPIGSPSPPPPVSVSPPPPVKSPPPPA---PVGSPPPPEKSPPPPA 570

Query: 61 PASSAPGHVAKLPQKEPVGCSKG---GGPPREDVGAP--LVTPSLLQMVRLRSVGAPGG 114  
 P +S P V P P V PP V +P + +P V AP  
 Sbjct: 571 PVASPPPPVKSPPPTLVASPPPPVKSPPPPAPVASPPPPVKSPPPTPVASPPPPAPVA 630

Query: 115 APTPALGPSAPQKPLRRALSGRASVPVAP---SSGLHAAVRLKACSLAASEGLSSAQPNG 171  
 + P + P P+ SP P P S+ S+ +S + P  
 Sbjct: 631 SSPPPMKSPPPPTPVSSPPPEKSPPPPPAKSTPPPEEYPTPTSVKSSPPPEKSLP-- 688

Query: 172 PPEAEPRPPQSPASTASFIFSKGSRKLQLERPVSPETQADLQRLNVAELRSISEQRPPQA 231  
 PP P PP T SK P SPE + + V+ + PP A  
 Sbjct: 689 PPTLIPSPPPQEKPTPSTPSKP-----PSSPEKPS-PKEPVSSPPQTPKSSPPPA 739

Query: 232 PKKSPKAPPPVARKPSVGV--PPPASPSYPRAEPLTAPP 268  
 P SP P PV+ P++ PP+ S P PL++PP  
 Sbjct: 740 PVSSPP-PTPVSSPPALAPVSSPPSVKSSPPAPLSSPP 777

Score = 162 (24.3 bits), Expect = 1.3e-08, P = 1.3e-08  
 Identities = 76/272 (27%), Positives = 99/272 (36%)

Query: 2 ADFPPPEEAFSVASPEPAG-PSGSP-ELVSSPAASSSSATALQIQQPGSPDPPPPAPPAPA 60  
 A P P SPEP PS P P + S A PP P P +PPA +  
 Sbjct: 427 ASAPMPSPHTPPDVSPPELPEPSVPAPAPMPMPTPHSPPADDDYVPPTPPVPGKSPATS 486

Query: 61 PASSAPGHVAKLPQKEPVGCSKGGPPREDVGAPLVTPSLLQMVRLRSVGAPGGATP-- 118  
 P+ A P V S PP+ VG+P P V+ S AP G+P+P  
 Sbjct: 487 PSPQVQPPAASTPPPSLVKLS----PPQAPVGSPP--PPP---VKTSPPAPIGSPSPPP 536

Query: 119 ---ALGPSAPQK-PLRRALSGRASVPVAPSSGLHAAVRLKACSLAASEGLSSAQPNGPPE 174  
 + P P K P A G SP P S A S + + PP  
 Sbjct: 537 PVSVPSPPPVKSPPPPAPVG--SPPPPEKSPPPAPVASPPPPVKSPPPTLVASPPPP 594

Query: 175 AEPRPPQSPASTASFIFSKGSRKLQLERPVSPETQADLQRLNVAELRSISEQRPPQAPKK 234  
 + PP +P ++ + P P A + + PP P+K  
 Sbjct: 595 VKSPPPAPVASPPPPVKSPPPTPVASPPPPAPVASSPPPMKSPPPPTPVSSPPP-PEK 653

Query: 235 SPKAPPPVARKPSVGVPPPASPSYPRAEPLTAPPTNGLP 273  
 SP PPP P PP P+ P + + PP LP  
 Sbjct: 654 SPPPPPPAKSTP---PPEEYPTPTSVKSSPPPEKSLP 688

Score = 159 (23.9 bits), Expect = 2.8e-08, P = 2.8e-08  
 Identities = 77/264 (29%), Positives = 103/264 (39%)

Query: 5 PPPEEAFSVASPEPAGPSGSP-ELVSSPAASSSSATALQIQQPGSP--DPPAP--PAP 59  
 PPP V+SP P P SP P SS ++ PP +P PP P P P  
 Sbjct: 916 PPPA---MVSSP-PMTKSSPP---PVVVSPPTVKSSPPAPVSSPPATPKSSPPP 966

Query: 60 APASSAPGHVAKLPQKEPVGCSKGGPPREDVGAPLVTPSLLQMVRLRSVGAPGGATP 119  
 AP + P V P PV S P AP+ +P + V+ AP +P P  
 Sbjct: 967 APVNLPPPEVKSSPPPTPVSSPPAPKSSPPPPAPMSSPPPE-VKSPPPAPVSSPPPP 1024

Query: 120 LGPSAPQKPLRRALSG-RASVPVAPSSGLHAAVRLKACSLAASEG---LSSAQNGPPEA 175  
 + P P+ ++ P PAP S V+ S + S P P +  
 Sbjct: 1025 VKSPPPAPVSSPPPPVKSPPPPAPVSSPPPPVKSPPPPAPISSPPPPVKSPPPPAPVSS 1084

Query: 176 EPRPPQSPASTASFIFSKGSRKLQLERPVSPETQADLQRLNVAELRSISEQRPPQAPKKS 235  
 P P +SP A S ++ P P A + A ++ S PP AP S  
 Sbjct: 1085 PPPPVKSPPPPAPV---SSPPPIKSPPPP---APVSSPPAPVKKPS--LPPAPVSS 1135

Query: 236 PK--APPPVARKPSVGVPPPA-SPSYRAEPLTAPP 268  
 P P +K +PPA S P + PP  
 Sbjct: 1136 PPPVTPAPPKKEQSLPPPAESQPPPSFNDIILPP 1171

Score = 143 (21.5 bits), Expect = 1.8e-06, P = 1.8e-06  
 Identities = 59/179 (32%), Positives = 77/179 (43%)

Query: 3 DFPPEEAFSSVASPEP-AGPSGSELVSSPAASSSSATA-LQIOPPGSP--DPPP--A 55  
 + PPPE S P P + P +P+ PA SS ++ PP +P PPP +  
 Sbjet: 970 NLPPPEVK--SSPPPTVSSPPAPKSSPPAPMSSPPPEVKSPPPAPVSSPPPPVK 1027

Query: 56 PPAPAPASSAPGHVAKLPQKEPVGCSKGGPPREDVGAPLVTPLSLQMVRLRSVGAPGGA 115  
 PP PAP SS P V P PV PP + P S V+ AP +  
 Sbjet: 1028 PPPAPVSSPPPPVKSPPPAPVSSPP--PPVKSPPPAPISSPPPPVKSPPPAPVSS 1084

Query: 116 PTPALGPSAPQKPLRRALSG-RASVPAPSSGLHAAVRLKACSLAASEGLSSAQNGPPE 174  
 P P + P P+ ++ P PAP S A +K SL +SS P PP  
 Sbjet: 1085 PPPPVKSPPPAPVSSPPPIKSPPPAPVSSPPAP-VKPPSLPPAPVSS--P--PPV 1139

Query: 175 AEP RPPQ 181  
 P PP+  
 Sbjet: 1140 VTPAPPK 1146

Score = 133 (20.0 bits), Expect = 2.3e-05, P = 2.3e-05  
 Identities = 50/132 (37%), Positives = 59/132 (44%)

Query: 1 MADFPPEEAFSSVASPEPAGP-SGSELVSSP--AASSSSATALQIOPPGSP--DPPP 54  
 M+ PPPE V SP P P S P V SP A SS ++ PP +P PPP  
 Sbjet: 1001 MSSPPPE----VKSPPPPAPVSSPPPPVKSPPPAPVSSPPPPVKSPPPAPVSSPP 1055

Query: 55 --APPAPAPASSAPGHVAKLPQKEPVGCSK--GGPPREDVGAPLVTPLSLQMVRLRS 108  
 +PP PAP SS P V P PV PP V +P P +  
 Sbjet: 1056 PVKSPPPPAPISSPPPPVKSPPPAPVSSPPPPVKSPPPAPVSSP--PPPIKSPPPAP 1113

Query: 109 VGAPGGAPT--PALGPSAP 125  
 V +P AP P+L P AP  
 Sbjet: 1114 VSSPPAPVKPPSLPPAP 1132

Score = 110 (16.5 bits), Expect = 8.0e-03, P = 8.0e-03  
 Identities = 41/121 (33%), Positives = 49/121 (40%)

Query: 5 PPPEEAFSS--VASPEPAGP-SGSELVSSP--AASSSSATALQIOPPGSP--DPPP 54  
 PPP S V SP P P S P V SP A SS ++ PP +P PPP  
 Sbjet: 1060 PPPPAPISSPPPPVKSPPPAPVSSPPPPVKSPPPAPVSSPPPIKSPPPAPVSSPP 1119

Query: 55 AP-----PAPAPASSAPGHVAKLPQKEPVGCSKGGPPREDVGAPLVTPLSLQMVRLRS 108  
 AP P PAP SS P V P K+ + PP E P +L +  
 Sbjet: 1120 APVKPPSLPPAPVSSPPPVTPAPPKKE---EQSLPPPAESQPPSFNDIILPIMANK 1176

Query: 109 VGAP 112  
 +P  
 Sbjet: 1177 YASP 1180

Score = 108 (16.2 bits), Expect = 1.3e-02, P = 1.3e-02  
 Identities = 46/155 (29%), Positives = 67/155 (43%)

Query: 114 GAPTALGPSAPQKPLRRALSGRASVPAPSSGLHAAVR-LKACS-LAASEGLSSAQNG 171  
 G PTP GP + P + A S +P+P+ + L S + A + P+  
 Sbjet: 408 GYPTGGGPPSSPVGKPAAS--APMPSPHTPPDVSPELPEPSVPVAPAPMPMPTPHS 464

Query: 172 PPEAEPRPPQSPASTASFIFSKGRKLQLERPVSPEQ---ADLQRNLVAELRSISEQR 227  
 PP + PP P S + S ++Q +P + Q + + +  
 Sbjet: 465 PPADDYVPPTPPVPGKSPATSPSPQVPPAASTPPPSLVKLSPPQAPVGSPPPPVKTTS 524

Query: 228 PPQAPKSKKAPPPVARKPSVGVPPPASPSYPRAEPLTAPP 268  
 PP AP SP PPV SV PPP S P P+ +PP  
 Sbjet: 525 PP-APIGSPSPPPV---SVVSPPPVKSPPPAPVGSPP 560

Pedant information for DKFZphmcf1\_lc23, frame 1

Report for DKFZphmcf1\_lc23.1

[LENGTH] 311  
 [MW] 31534.58  
 [pI] 9.48  
 [KW] All\_Alpha  
 [KW] LOW\_COMPLEXITY 38.59 %

SEQ MADFPPEEAFSSVASPEPAGPSGSELVSSPAASSSSATALQIOPPGSPDPPAPAPAPA  
 SEG .....XX  
 PRD ccc

SEQ PASSAPGHVAKLPQKEPVGCSKGGPPREDVGAPLVTPLSLQMVRLRSVGAPGGAPTAL  
 SEG xxxxxx.....XX



```
PRD      ccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc
SEQ      GPSAPQKPLRRALSGRASPVPAPSSGLHAAVRLKACSLAASEGLSSAQPNGPPEAEPRPP
SEG      xxxxx.....xxxxxxxxxxxxxx
PRD      cccccchhhhhhhhhccccccccchhhhhhhhhhhhhhhccccccccccccccccccccc

SEQ      QSPASTASFIFSKGSRKLQLERPVPETQADLQRNLVAELRSISEQRPPQAPKKSPPKAPP
SEG      xxxxx.....xxxxxxxxxxxxxx
PRD      cccccceeeccccChhhhccccccccchhhhhhhhhhhhhhhccccccccccccccccccccc

SEQ      PVARKPSVGVPASPSPYPRAEPLTAPPTNGLPHTQDRTKRELAENGVLQLVGPEEKMG
SEG      xxxxxxxxxxxxxxxxxxxxxxxxx.....
PRD      cccccccccccccccccccccccccccccccccccccccccchhhhhhhcccccecccccccccc

SEQ      LPGSDSQKELA
SEG      .....
PRD      ccccccccccc
```

(No Prosite data available for DKFZphmcf1\_1c23.1)

(No Pfam data available for DKFZphmcf1\_1c23.1)

DKFZphmcfl\_1e15

group: transmembrane protein

DKFZphmcfl\_1e15 encodes a novel 454 amino acid protein with similarity to C. elegans proteins and transporter proteins.

The novel protein is similar to the PTR2 family of proton/oligopeptide symporter proteins and the D-xylose-proton symporter. Thus, the protein is a transporter of a so far unknown compound.

The new protein can find application as a new transporter in eukaryotic cells, e.g. in drug transport into cells.

similarity to D-XYLOSE TRANSPORTER  
membrane regions: 9

complete cDNA, complete cds, EST hits  
matches cDNA encoding cell growth inhibiting factor (E12646)

Sequenced by DKFZ

Locus: unknown

Insert length: 1957 bp

Poly A stretch at pos. 1947, polyadenylation signal at pos. 1929

```

1  GGTGCAGCGC  CCGGGCTGAG  CGACAGCAAG  TGCAGCGGGC  TCCTACCCCG
51  GGTGAGGGGT  GGCCTCCGCG  TGGGATCGTG  CCCTCTTCAG  CCCGCTCCTG
101  TCCTCCGACAT  CACGTGTATT  CCGCACGTCC  CCTCCGCGCT  GTGTGTCTAC
151  TGAGACGGGG  AGCGGTGACA  GGGCCCGGGT  CCCTCTCAG  TGGTGTCTG
201  TGCTTCAGGG  CAAGCTCCCC  GTCTCCGGGC  GCACTTCCCT  CGCTGTGTG
251  CGGTCCATCC  TCCTTTCTCC  AGCCTCCTCC  CCTCGCAGGT  GGGATCGTCG
301  GTGGGACCGG  AGCGGGGGCG  GCGCGGGCCC  CCCGGGACCA  TGGCGGGGTC
351  CGACACCGCG  CCCTTCCTCA  GCCAGGCGGA  TGACCGGGAC  GACGGGCGAG
401  TGCTTGGCAC  CCGGGGGTGG  CCAGGGTCCA  CGGGGAACCC  GAAGTCCGAG
451  GAGCCCGAGG  TCCCGGACCA  GGAGGGGCTG  CAGCGCATCA  CCGGCCTGTC
501  TCCCGGCGGT  TCGGCTCTCA  TAGTGGCGGT  GCTGTGTACT  ATCAATCTCC
551  TGAATACAT  GGACCGCTTC  ACCGTGGCTG  TGTTTATCTC  CAGTTACATG
601  GTGTTGGCAC  CTGTGTTTGG  CTACCTGGGT  GACAGGTACA  ATCGGAAGTA
651  TCTCATGTGC  GGGGGCATTG  CCTTCTGGTC  CCTGCTGACA  CTGGGGTCAT
701  CCTTCATCCC  CGGAGAGCAT  TTCTGGCTGC  TCCTCCTGAC  CCGGGGCTG
751  GTGGGGTTCG  GGGAGGCCAG  TTATTCCACC  ATCGCGCCCA  CTCTCATTGC
801  CGACCTCTTT  GTGGCCGACC  AGCGGAGCCG  GATGCTCAGC  ATCTTCTACT
851  TTGCCATTCC  GGTGGGCACT  GGTCTGGGCT  ACATTGCAGG  CTCGAAAGTG
901  AAGGATATGG  CTGGAGACTG  GCACTGGGCT  CTGAGGGTGA  CACCGGGTCT
951  AAGGATGGTG  GCGGTCTGCG  TGCTGTTCCT  GGTAGTCCGG  GAGCGGCCAA
1001  GGGGAGCCGT  GGAGCGCCAC  TCAGATTTCG  CACCCCTGAA  CCCCACCTCG
1051  TGGTGGGCAG  ATCTGAGGGC  TCTGGCAAGA  AATCTCATCT  TTGGACTCAT
1101  CACCTGCCTG  ACCGGAGTCC  TGGGTGTGGG  CCTGGGTGTG  GAGATCAGCC
1151  GCGGGCTCCG  CCACTCCAAC  CCGGGGCTG  ATCCCTGGT  CTGTGCCACT
1201  GGCTCCTG  GCTCTGCACC  CTCTCTCTTC  CTGTCCCTTG  CTGCGCCCG
1251  TGGTAGCATC  GTGGCCACTT  ATATTTTCAT  CTTTATTGGA  GAGACCTCC
1301  TGTCATGAA  CTGGGCCATC  GTGGCCGACA  TTCTGTGTGA  CTGTGTGATC
1351  CCTACCCGAC  GCTCCACCGC  CGAGGCCCTC  CAGATCGTGC  TGTCCACCT
1401  GCTGGGTGAT  GCTGGGAGCC  CCTACCTCAT  TGGCCTGATC  TGTACCCGCC
1451  TGCGCCGGAA  CTGGCCCCC  TCCTTCTTGT  CCGAGTTCCG  GGCTCTGCAG
1501  TTCTCGCTCA  TGCTCTGCGC  GTTTGTTGGG  GCACTGGGCG  GCGCAGCCTT
1551  CCTGGGCACC  GCCATCTTCA  TTGAGGCCGA  CCGCGGGCGG  GCACAGCTGC
1601  ACGTGCAGGG  CTGTCTGCAC  GAAGCAGGGT  CCACAGACGA  CCGGATTGTG
1651  GTGCCCCAGC  GGGGCCGCTC  CACCCGCGTG  CCCGTGGCCA  GTGTGCTCAT
1701  CTGAGAGGCT  GCCGCTCACC  TACCTGCACA  TCTGCCACAG  CTGGCCCTGG
1751  GCGCACCCCA  CGAAGGGCCT  GGGCCTAACC  CCTTGGCCCTG  GCGCAGCTTC
1801  CAGAGGGACC  CTGGGCGGTG  TGCCAGCTCC  CAGACACTAC  ATGGGTAGCT
1851  CAGGGGAGGA  GGTGGGGGTG  CAGGAGGGGG  ATCCCTCTCC  ACAGGGGCAG
1901  CCGCAAGGGC  TCGGTGCTAT  TTGTAACGGA  ATAAAATTG  TAGCCAGAAA
1951  AAAAAAA

```

## BLAST Results

-----

Entry E12646 from database EMBL:  
cDNA encoding cell growth inhibiting factor.  
Score = 3046, P = 2.2e-131, identities = 640/659

## Medline entries

No Medline entry

## Peptide information for frame 1

ORF from 340 bp to 1701 bp; peptide length: 454  
Category: similarity to known protein

```
1 MAGSDTAPFL SQADDPDDGP VPGTPGLPGS TGNPKSEEP VPDQEGLRQI
51 TGLSPGRSAL IVAVLCYINL LNYMDRFTVA VFISYMYLA PVFGYLGDRY
101 NRKYL MCGGI AFWSLVTLGS SFIPGEHFWL LLLTRGLVGV GEASYSTIAP
151 TLIADLFVAD QRSRMLSIYF FAIPVGSGLG YIAGSKVKDM AGDWHWALRV
201 TPGLGVVAVL LFLVVRPEP RGAVERHSDL PPLNPTSWMA DLRALARNLI
251 FGLITCLTGV LGVGLGVEIS RRLRHSNPRA DPLVCATGLL GSAPFLFLSL
301 ACARGSIVAT YIFIFIGETL LSMNWAIVAD ILLYVVIPT RSTAEAFQIV
351 LSHLLGDAGS PYLIGLISDR LRRNWPPSFL SEFRALQFSL MLCFAVFGALG
401 GAAFLGTAIF IEADRRRAQL HVQGLLHEAG STDDRIVVPQ RGRSTRVPVA
451 SVLI
```

## BLASTP hits

No BLASTP hits available

## Alert BLASTP hits for DKFZphmcf1\_1e15, frame 1

TREMBL:CEC13C4\_1 gene: "C13C4.5"; Caenorhabditis elegans cosmid C13C4,  
N = 3, Score = 441, P = 5.2e-76

TREMBL:CEC39E9\_10 gene: "C39E9.10"; Caenorhabditis elegans cosmid  
C39E9, N = 2, Score = 449, P = 8.2e-69

TREMBL:CEF09A5\_1 gene: "F09A5.1"; Caenorhabditis elegans cosmid F09A5,  
N = 3, Score = 413, P = 9.1e-60

TREMBL:ATF6H11\_18 gene: "F6H11.180"; product: "predicted protein";  
Arabidopsis thaliana DNA chromosome 5, BAC clone F6H11 (ESSAII  
project), N = 3, Score = 193, P = 2.5e-24

SWISSPROT:XYLT\_LACBR D-XYLOSE-PROTON SYMPORT (D-XYLOSE TRANSPORTER)., N  
= 1, Score = 180, P = 7.9e-11

>TREMBL:CEC39E9\_10 gene: "C39E9.10"; Caenorhabditis elegans cosmid C39E9  
Length = 488

## HSPs:

Score = 449 (67.4 bits), Expect = 8.2e-69, Sum P(2) = 8.2e-69  
Identities = 88/204 (43%), Positives = 125/204 (61%)

Query: 58 SALIVAVLCYINLLNYMDRFTVAVFISSYMYLAPVFGYLGDRYNRKYLMCGGIAFWSLVT 117  
+ ++ V Y N+ + + VF+ S+MV +PV GYLGD+NRK++M G+ W  
Sbjct: 29 AGVLTVQVQTYYNISDSLGGLIQTVFLISFMVFSVPCGYLGDRFNKRWIMIIIGVIGLAV 88

Query: 118 LGSSFIPGEHFWL LLLTRGLVGVGEASYSTIAPT LIADLFVADQRSRMLSIYF FAIPVGS 177  
LGSSF+P HFWL L+ R VG+GEASYS +AP+LI+D+F +RS + IFYFAIPVGS  
Sbjct: 89 LGSSFVPANHFVLFVLRSEFVGIGEASYSNVAPSLISDMFNGQKRSTVFMIFYFAIPVGS 148

Query: 178 GLGYIAGSKVKDMAGDWHWALRVTPGLGVVAVLLFLVVRPEPRGAVER----HSDLPL 233  
GLG+I GS V + G W W +RV+ G++ ++ L L EP RGA ++ D+  
Sbjct: 149 GLGFIVGSNVATLTGHMQGWIRVSAIAGLIVMIALVLFTYPEPERGAADKAMGESKDVVVT 208

Query: 234 NPTSWWADLRALARNLIFGLITCLTG 259  
T++ DL L + L+ C G  
Sbjct: 209 TNTTYLEDLVILLKTPT--LVACTWG 232

Score = 267 (40.1 bits), Expect = 8.2e-69, Sum P(2) = 8.2e-69  
Identities = 74/212 (34%), Positives = 113/212 (53%)

Query: 249 LIFGLITCLTGLVGLGVEISRRL-----RHSNPRADPLVCATGLLGSAPFLFLSL 300  
L FG IT G++GV G +S+ L R RA PLV G L +APFL + +  
Sbjct: 277 LYFGAITTAGGLIGVIFGSMLSKWLVAGWGPFRRLQTDRAQPLVAGGALLAAPFLIGM 336

Query: 301 ACARGSIVATYIFIFIGETLLSMNWAIVADILLYVVIPTRRSTAEAFQIVLSHLLGDAGS 360

```

SV YI IF G T + NW + D+L V+ P RRSTA ++ +++SHL GDA
Sbjct: 337 IFGDKSLVLLYIMIFGFIGTIFMCFNWLGNIDMLTTVIHPNRRSTAFSYFVLVSHLFGDASG 396

Query: 361 PYLIGLISDLRLRN--WPPSFLSEFRALQFSMLCAVFVGLGGAALFGTAIFIEDRR-- 416
      PYLIGLISD +R P ++ +L + C + L + +++ + +DR+
Sbjct: 397 PYLIGLISDAIRHGSTYPKD--QYHSLVSATYCCVALLLSAGLYFVSSLTIVSDRKKF 453

Query: 417 RAQLHVQGLLHEA--GSTD--DRVVPQGRGRSTRV 447
      RA++ + L + STD +RI + S +R+
Sbjct: 454 RAEMGLDDLQSKPIRSTDSLERIGINDDVASSRL 488

Score = 70 (10.5 bits), Expect = 5.9e-24, Sum P(2) = 5.9e-24
Identities = 25/89 (28%), Positives = 41/89 (46%)

Query: 62 VAVLCYINLLNMYMDRFTVAVFISSYMVLPVFGYLGDRYNRKYLMCGGIAFWSLVT--LG 119
      V L +NLLNY+DR+TVA ++ + LG +L+ +S V LG
Sbjct: 11 VTLFVVLNLLNRYDRTYVAGVLTVQVQTYNIDSLSGLGIQTVFLLI--SFMVFPVCGYL 68

Query: 120 SSFIPGEHFVLLLLTRGLVGVGEASYSTIAP 150
      F W++++ G +G S+ P
Sbjct: 69 DRF--NRKWMIMIGVG-IWLGVAVLGSSFP 95

```

Pedant information for DKFZphmcfl\_1el5, frame 1

## Report for DKFZphmcf1 1e15.1

```

[LENGTH]      454
[MW]           49013.35
[pI]           7.66
[HOMOL]       TREMBL:CEC13C4_1 gene: "C13C4.5"; Caenorhabditis elegans cosmid C13C4 2e-51

[BLOCKS]       BL01022D
[PROSITE]      MYRISTYL      11
[PROSITE]      CAMP_PHOSPHO_SITE      1
[PROSITE]      CK2_PHOSPHO_SITE      3
[PROSITE]      PROKAR_LIPOPROTEIN      1
[PROSITE]      GLYCOSAMINOGLYCAN      1
[PROSITE]      PKC_PHOSPHO_SITE      4
[KW]           TRANSMEMBRANE      8
[KW]           LOW COMPLEXITY      15.42 %

```

[illegible]

**PCT/IB00/01496**

Prosites for DKFZphmcfl\_1e15.1

(No Pfam data available for DKFZphmcf1\_1e15.1)

DKF2phmcf1\_lg13

-----

group: mammary carcinoma derived

DKF2phmcf1\_lg13 encodes a novel 573 amino acid protein with very weak similarity to the human KIAA0543 protein and Musca domestica hermes transposase.

No informative BLAST results; No predictive prosite, pfam or SCOP motif.

The new protein can find application in studying the expression profile of mammary carcinoma-specific genes.

similarity to KIAA0766

complete cDNA, complete cds, few EST hits  
on genomic level encoded by AC005020, no splicing, genomic?

Sequenced by DKF2

Locus: unknown

Insert length: 2210 bp

Poly A stretch at pos. 2200, polyadenylation signal at pos. 2176

```

1  GAAACCTGAT CTCATAAAC CTAGGTCACA AAGGACAGCC CTGCAAAACA
51  GACCCATATT GGATCAAGTG AGCCAGTTCC TGGAACTGA ATAATGACTC
101 CTGAATCAAG GGATACTACA GATTGTCTC CAGGGGGTAC CCAGGAGATG
151 GAAGGCATCG TGATAGTGAA GGTGGAGGAG GAAGATGAAG AAGACCATT
201 TCAAAAGGAA AGAAACAAAG TAGAGTCATC GCCACAAGTT CTCAGTCGCT
251 CTACAACAT GAATGAGAGA GCCTTATTGT CATCGTATT AGTTGCATAT
301 AGAGTGGCAA AAGAGAAAAT GGCTCAGACA GCGGCTGAAA AAATATACCT
351 TCCAGCATGT ATGGACATGG TACGGACAAT TTTTGATGAC AAATCAGCTG
401 ATAAACTAAG AACTATACCT CTTAGTGATA ATACAATATC TCGTCGAATC
451 TGTACGATTG CAAAACATTG GGAAGCAATG CTTATTACAC GGCTGCAGTC
501 CGGTATAGAC TTTGCAATCC AACTCGATGA GAGCACTGAT ATTGCAAGTT
551 GTCCACACTC CTTGGTTTAT GTCAGATATG TGTGGCAAGA TGATTTTGTA
601 GAGGATCTCT TATGTTGTTT AAATTTAAAT TCACATATAA CTGGATTAGA
651 TTTATTTACT GAATTAGAAA ACTGCCTTCT TGGTCAGTAT AAATTAACCT
701 GGAAACATTG TAAAGGAATT TCAAGTGATG GAACAGCAAA TATGACCGGA
751 AAACACAGCA GACTTACTGA AAAATTGTTA GAAGCAACCC ACAACATGTC
801 TGTTTGGAAAT CACTGTTTTA TTCATCGAGA AGCTTTGGTA TCCAAAGAAA
851 TTTCACCAAG TCTGATGGAT GTATTGAAA ATGCAGTGAA AACTGTTAAT
901 TTTATTAAAG GAAGCTCACT GAATAGCCGA CTTCTCGAAA TATTTTGTTT
951 AGAGATTGGA GTGAACCACA CCCACTTATT GTTTCATACA GAAGTTCGTT
1001 GGCTTTCTCA AGGAAAAGTA TTGAGCAGAG TATATGAAC CAGGAACGAG
1051 ATTTACATTT TTCTCGTTGA AAAGCAATCT CATTGGCAA ATATTTTGTA
1101 AGACGACATT TGGGTAACAA AATTGGCATA TTTAAGTGAT ATTTTGGCAA
1151 TTCTTAATGA ATTAAGCCTG AAAATGCAGG GGAACAAACA TGATATATTT
1201 CAGTATCTTG AACATATTCT AGGATTCCAA AAGACGTTAT TATTGTGGCA
1251 AGCAAGACTT AAAAGTAACC GCCCTAGCTA CTATATGTTT CCAACATTAT
1301 TGCAACACAT CGAAGAGAAC ATTATTAATG AAGACTGCTT AAAAGAAATA
1351 AAATTAGAGA TATTGTTGCA TCTCACTTCT TTGCTCAAAA CTTTAAATTA
1401 TTACTTTCCG GAAGAGAAAT TTGAATCATT AAAGGAAAT ATTTGGATGA
1451 AAGATCCATT TGCTTTTCAA AACCCAGAAAT CAATAATTGA GTTAAACTTG
1501 GAGCCTGAAG AAGAGAAATG ATTATTGCAG CTCAGTTCAT CATTACACT
1551 AAAGAATTAT TATAAGATAT TAAGTTTATC AGCATTTTGG ATTAAGATTA
1601 AAGATGACTT TCCACTGCTA AGTAGGAAGA GTATATTGCT GTTACTACCA
1651 TTCACAACAT CATATTGTGT TGAAGTAGGA TTTTCAATCT TGACACGGTT
1701 AAAACAACAA AAGAGAAATA GGCTCAATAG TGCACAGAT ATGCGGGTAG
1751 CATTATCTTC ATGTGTTTCT GACTGGAAGG AACTTATGAA CAGACAAGCA
1801 CACCCATCAC ATTAATATCA AACTTTACAA AATTCTGTGT ATAGCCAGGT
1851 GTGGTGGCTT ACGCCTGTAA TCCCAGCAGT GGGAGACCGA GGTGGGCAGA
1901 TCACTTGAGT TCAAGACCAG CCTGGCCAAC ATGGTGAAC CCCATCTCTA
1951 CTAAAAAATG AACCTTAGC CAGGCGTGGT GGCACATGCC TGCAGTCCCA
2001 GTTACTTGGG TGCTGAGGC AGGAGAATCT CTTAAACCAG GAAGGCAGAG
2051 ATTGCAGTGA GCTGAGATAA TCCCAGTCCA TTCCAGCCTG GGAACAGCG
2101 TGAGACTTCA TCTCAAAAAA AAAAATTGT ATTTGACTT TTAAGGGAT
2151 TTTGCAGTAT GTTGATGTTA AACGTTAATA AAATTATATT TGAATTAGG
2201 AAAAAAAAAA

```

## BLAST Results

-----

Entry AC005020 from database EMBL:  
Homo sapiens clone GS259H13; HTGS phase 1, 4 unordered pieces.  
Score = 9110, P = 0.0e+00, identities = 1822/1822

## Medline entries

No Medline entry

## Peptide information for frame 1

ORF from 94 bp to 1812 bp; peptide length: 573  
 Category: similarity to unknown protein

```

1 MTPESRDITD LSPGGTQEME GIVIVKVEEE DEEDHFQKER NKVESSPQVL
51 SRSTTMNERA LLSSYLVAIR VAKKMAHTA AEKIILPACM DMVRTIFDDK
101 SADKLRTIPL SDNTISRRIC TIAKHLEAML ITRLQSGIDF AIQLDESTDI
151 ASCPTLLVYV RYVWQDDFVE DLLCCLNLNS HITGLDLFTE LENCLLGQYK
201 LNWKHKCGIS SDGTANMTGK HSRLTEKLE ATHNNVWNN CFHREALVS
251 KEISPSLMDV LKNAVKTVMF IKGSSLNSRL LEIFCSEIGV NHTHLLFHTF
301 VRWLSQGVKL SRVYELRNEI YIFLVEKQSH LANIFEDDIW VTKLAYLSDI
351 FGILNELSLK MQGKNNDIFQ YLEHILGFK TLLMQARLK SNRPSYMF
401 TLLQHIEENI INEDCLKEIK LEILLHLTSL SQTFNYYFPE EKFECLKENI
451 WMKDPFAFQN PESITELNLE PEEENELLQL SSSFTLKNNY KILSLSAFWI
501 KIKDDFPLLS RKSILLLLPF TTYLCELGF SILTRLTKK RNRNSAPDM
551 RVALSSCVPD WKELMNRQAH PSH
  
```

## BLASTP hits

Entry AC004877\_3 from database TREMBLNEW:  
 gene: "WUGSC:H\_DJ0751H13.2"; product: "KIAA0543 protein"; Homo sapiens  
 PAC clone DJ0751H13 from 7q35-qter, complete sequence.  
 Score = 86, P = 4.4e-03, identities = 46/179, positives = 78/179

Entry MD36211\_1 from database TREMBL:  
 product: "Hermes transposase"; Musca domestica Hermes transposase  
 gene, complete cds.  
 Score = 105, P = 3.0e-02, identities = 101/465, positives = 202/465

## Alert BLASTP hits for DKFZphmcf1\_lg13, frame 1

TREMBL:AB018309.1 gene: "KIAA0766"; product: "KIAA0766 protein"; Homo sapiens mRNA for KIAA0766 protein, complete cds., N = 1, Score = 300, P = 1.1e-23

>TREMBL:AB018309.1 gene: "KIAA0766"; product: "KIAA0766 protein"; Homo sapiens mRNA for KIAA0766 protein, complete cds.  
 Length = 607

## HSPs:

Score = 300 (45.0 bits), Expect = 1.1e-23, P = 1.1e-23  
 Identities = 120/485 (24%), Positives = 229/485 (47%)

```

Query:      89 CMD-MVRTIFDDKSADKLRTIPLSDNTISRRICTIAKHLEAMLITRLQSGIDFAIQLDES 147
             CM+ ++R + + L + LS + +RI +I ++L L R + +++ LD+
Sbjct:     124 CHEVLLREVLPFH-VSVLQGVOLSPDITRQILSIDRNLRLNQLFNRRDFKAYSLALDDQ 182

Query:     148 TDIASCPTLLVYRYVWQD-DFVEDLLCCLNLNSHIT-GLDLFTELENCLLGQYKLNWKH 205
             +A LLV++R V + + EDLL +NL H + G + LE+ L L + +
Sbjct:     183 AFVAYENYLLVFIRGVGPELEVQEDLLTIINLTHFVSGALMSAILES--LQTAGLSLQR 240

Query:     206 KGISSDGTANMTGKHSRLTEKLEATHNNVWNN--HC--FIHREALVSKEISPSLMDVL 261
             G+++ T M G++S L + E + WN H F+H E L S ++ + ++
Sbjct:     241 MVGLTTTHTLRMIGENSGLVSYMREKAVSPNCWNVIHYSGLHLELLSSYOVDVN--QII 298

Query:     262 KNAVKTVMFIKGSLSRLLEIFCSEIGVNNHTHLLFHTFVR-WLSQGVLSRVYELANEI 320
             + + IK + + +E H + + WL +GK L ++ LR E+
Sbjct:     299 NTISEWIVLIKTRGVRRPEFQTLTTESESEHGERVNGRCLNWLRRGKTLKLIFSLRKEM 358

Query:     321 YIFLVEKQSHLANIFEDDIWVTKLAYLSDFGILNELSLKMQGKNNDIFQYLEHILGFK 380
             FLV + + F D W+ +L DI L ELS +++ +HI F+
Sbjct:     359 EAFLVSVGATTVH-FSDKQWLCDFGLVDIMEHLRELSEELRVSKVFAAAAFDICTFEV 417
  
```

Query: 381 TLLLQWARKLSNRPSYMFPTLLQHIEE----NIINEDCKEIKLEILLHLTSLSQTFNY 436  
L L+Q ++ + + FP L + + + E N + E + + + + L + F

Sbjct: 418 KNLNLFQRHIEEKNLTLD--FPALREVVDLKQNKQEKEDKIFDPDRYQMVI---CRLQKEFER 473

Query: 437 YFPEEFESLKENIWM-KDPPAFQNPSIEIENLNPEEEENELLQLSSSFTLKNYYKILSL 495  
F + + F + K + + + + + PF F + + I + + E L + L + + L N Y + I L

Sbjct: 474 HFKDLRF--IKKDLELFSNFNFNFKPEYAPISVRVE-----LTKLQANTNLWNEYRIKDL 525

Query: 496 SAFWIKIK-DFPFLSRKSILLLPFTTTYLCELGFSILTRLTKKRNLNSA---PDMR 551  
G F + + + + P + + + + + F + + + CE FS LTR + L R

Sbjct: 526 GQFYAGLSAESYPIIKGVACKVASLFDNSQICEKAFSYLTRNQHTLSQPLTDEHLQALFR 585

Query: 552 VALSSCPVMDKELMNRQAHPSH 573  
VA + P W + L + R + S +

Sbjct: 586 VATTEMEPGWDVLV-RERNSEN 606

Score = 290 (43.5 bits), Expect = 1.5e-22, P = 1.5e-22  
Identities = 120/485 (24%), Positives = 228/485 (47%)

Query: 89 CHD-MVRTIFDDKSADKLRTIP LSDNTISRRICTIAKHLAMLITRLQSGIDFAIQLODES 147  
CM+ ++R + + + L + + LS + + RI + I + L L R + + + ++ LD+

Sbjct: 124 CHEVLLREVLPLEPH-VSVLGQVDLSPDITRQRILSIDRNLRNQLFNRRADFKAYSALADDQ 182

Query: 148 TDIASCEPTLVVVRYVQD-DFVEDDLCCCLNLSHIT-GDLFTLELNCGLQQYKLNWKH 205  
+ A LLV++R V + + EDLL + NL H + + LE+ L L + +

Sbjct: 183 AFVAYENYLLVFIRGVGPPELVQEDELTIINLTHHFSVGALMSAILES--LQTAGLSLQR 240

Query: 206 CKGISGDGTANMTKGHSRLTEKLLEATHNNNAVNHCFIHREALVSKEISPSLMVD-LKNA 264  
G++ T M G++S L L + + E + WN H + + E + S D V +

Sbjct: 241 MVGLTTTHTLRMIGENSGLVSYMREKAVSPCNWC--VIHYSGLFHLELLSSY-DVDVNQOI 297

Query: 265 VKTVN----FIKGSLSNRLLEIFCSIEGVNHTHLHFTEVR-WLSQGKVLSRVLYELNE 319  
+ T + + IK + + + + + E H + + + WL + GK L + + LR E

Sbjct: 298 INTISEWIVLIKTRGVARRPEFQTLTTESESEHGERVERNGRCINNWLRRGKTLKLI FSLRKE 357

Query: 320 IYIFLVEKQSHLANIFEDDIWWTKLAYLSIDFGILNELSLMKMGKMNIDIFYEHLEGFQ 379  
+ FLV + + F D W + + L DI L EL S + + + + + HI F +

Sbjct: 358 MEAFLVSVGGATTVH-FSDQKWLCDFGLVDIMEHRELSELRSVKVFAAAAFDHICTFE 416

Query: 380 KTLMLQARLKS NRPSYMFPTLLQHIEENIINEDCKEIKL----EILLHLTSLSQTFN 435  
L L+Q ++ + + FP L + + + E + + + K + + + + L + F

Sbjct: 417 VKNLNLFQRHIEEKNLTLD--FPALREVVD--LKQNKQEKEDKIFDPDRYQMVICRLQKEFE 472

Query: 436 YFPEEFESLKENIWM-KDPPAFQNPSIEIENLNPEEEENELLQLSSSFTLKNYYKILSL 494  
+ HF + F + + K + + + + + PF F + + I + + E L + L + + L N Y + I

Sbjct: 473 RHFKDLRF--IKKDLELFSNFNFNFKPEYAPISVRVE-----LTKLQANTNLWNEYRIKDL 524

Query: 495 LSAFWIKIK-DFPFLSRKSILLLPFTTTYLCELGFSILTRLTKKRNLNSA---PDM 550  
L F + + + + P + + + + + F + + CE FS LTR + L R

Sbjct: 525 LGQFYAGLSAESYPIIKGVACKVASLFDNSQICEKAFSYLTRNQHTLSQPLTDEHLQALF 584

Query: 551 RVALLSSCPVMDKELMNRQAHPSH 573  
RVA + P W + L + R + S +

Sbjct: 585 RVATTEMEPGWDVLV-RERNSEN 606

Pedant information for DKFZphmcf1\_lg13, frame 1

Report for DKFZphmcf1\_1g13.1

```

[LENGTH]      573
[MW]           66276.85
[pI]           5.82
[HOMOL]        TREMBL:AB018309.1 gene: "KIAA0766"; product: "KIAA0766 protein"; Homo sapiens
mRNA for KIAA0766 protein, complete cds. 1e-18
[PROSITE]      MYRISTYL      3
[PROSITE]      CK2_PHOSPHO_SITE      10
[PROSITE]      TYR_PHOSPHO_SITE      1
[PROSITE]      PKC_PHOSPHO_SITE      9
[PROSITE]      ASN_GLYCOSYLATION      2
[KW]           All_Alpha
[KW]           LOW_COMPLEXITY      8.90 %

SEQ      MTPESRDTTDLSPGGTQEMEGIVIVKVEEEDDEHDFQKERNKVESSPQVLSRSTTMNERA
SEG      .....xxxxxxxx.....
PRD      cccccccccccccccccceccccchhhhhhhhhhhccccceccccchhhh

SEQ      LLSSYLVAYRVAKEMAHTAAEKIILPACMDMVRTIFDDKSADKLRTIPLSDNTISRRIIC

```



[illegible]

Prosites for DKFZphmcf1\_1g13.1

PS000001	216-220	ASN GLYCOSYLATION	PD0C00001
PS000001	291-295	ASN GLYCOSYLATION	PD0C00001
PS000005	116-119	PKC_PHOSPHO_SITE	PD0C00005
PS000005	218-221	PKC_PHOSPHO_SITE	PD0C00005
PS000005	225-228	PKC_PHOSPHO_SITE	PD0C00005
PS000005	358-361	PKC_PHOSPHO_SITE	PD0C00005
PS000005	391-394	PKC_PHOSPHO_SITE	PD0C00005
PS000005	445-448	PKC_PHOSPHO_SITE	PD0C00005
PS000005	485-488	PKC_PHOSPHO_SITE	PD0C00005
PS000005	510-513	PKC_PHOSPHO_SITE	PD0C00005
PS000005	538-541	PKC_PHOSPHO_SITE	PD0C00005
PS000006	55-59	CK2_PHOSPHO_SITE	PD0C00006
PS000006	79-83	CK2_PHOSPHO_SITE	PD0C00006
PS000006	95-99	CK2_PHOSPHO_SITE	PD0C00006
PS000006	136-140	CK2_PHOSPHO_SITE	PD0C00006
PS000006	183-187	CK2_PHOSPHO_SITE	PD0C00006
PS000006	189-193	CK2_PHOSPHO_SITE	PD0C00006
PS000006	256-260	CK2_PHOSPHO_SITE	PD0C00006
PS000006	445-449	CK2_PHOSPHO_SITE	PD0C00006
PS000006	463-467	CK2_PHOSPHO_SITE	PD0C00006
PS000006	546-550	CK2_PHOSPHO_SITE	PD0C00006
PS000007	364-372	TYR_PHOSPHO_SITE	PD0C00007
PS000008	137-143	MYRISTYL	PD0C00008
PS000008	273-279	MYRISTYL	PD0C00008
PS000008	289-295	MYRISTYL	PD0C00008

(No Pfam data available for DKFZphmcf1\_lgl3.1)

DKFZphtes3\_14q5

-----

group: testes derived

DKFZphtes3\_14q5 encodes a novel 379 amino acid protein with strong similarity to murine cell growth regulating nucleolar protein LYAR.

The novel protein is very similar to murine Ly-1 antibody reactive clone protein (LYAR). It contains a ATP/GTP-binding site motif A (P-loop, interacts with one of the phosphate groups of a ATP/GTP nucleotide), but not the zinc finger motif and nuclear localization signals of lyar.

No informative BLAST results: No predictive prosite, pfam or SCOP motive.

The new protein can find application in studying the expression profile of testis-specific genes.

strong similarity to cell growth regulating nucleolar protein LYAR, of mouse

complete cDNA, complete cds, EST hits

Sequenced by BMFZ

Locus: unknown

Insert length: 1503 bp

Poly A stretch at pos. 1467, polyadenylation signal at pos. 1440

```
1 CCCAGAGGTC CGACCTGGGA GGCTGGGGCT CAGAGAGCAA TGTTTGCTGT
51 CTTCCATTGG AGTGACTGAA TTTCTACATG ACGGCTTTTT GACAAGACTT
101 AAAACCTGTC TTGGATAGAG AATATTTAGC CATTACCTTA AAAATGGTAT
151 TTTTACATG CAATGCATGT GGTGAATCAG TGAAGAAAAT ACAAGTGGAA
201 AAGCATGTGT CTGTTTGCGA AAACGTGTGA TGCTTTTCTT GCATTGACTG
251 CGGTAAGAT TTCTGGGGCG ATGACTATAA AAACCCAGTG AAATGCATAA
301 GTGAAGATCA GAAGTATGGT GGCAAGGCT ATGAAGGTAA AACCCACAAA
351 GGCACATCA AACAGCAGGC GTGGATTGAG AAAATTAGTG AATTAATAAA
401 GAGACCCAAT GTCAGCCCA AAGTGAGAGA ACTTTAGAG CAAATTAGTG
451 CTTTTCACAA CGTTCCCGAG AAAAGGGCAA AATTCAGAA TTGGATGAAG
501 AACAGTTTAA AAGTTCATAA TGAATCCATT CTGGACCAGG TGTGGAATAT
551 CTTTCTGAA GCTTCCAACA GCGAACCATG CAATAAGGAA CAGGATCAAC
601 GGCCACTCCA CCCAGTGGCA AATCCACATG CAGAAATCTC CACCAAGGTT
651 CCAGCCTCCA AAGTGAAGA CGCCGTGGAA CAGCAAGGGG AGGTGAAGAA
701 GAATAAAGA GAAAGAAAGG AAGAACGGCA GAAGAAAAGG AAAAGAGAAA
751 AGAAGAACT AAAATTAGAA AACCACAGG AAAACTCAAG GAATCAGAAG
801 CCTAAGAAGC GCAAAAAGGG ACAGGAGGCT GACCTGAGG CTGGTGGGGA
851 GGAATGCCCT GAGGCCAATG GCTTCGAGG GAAGAGGAGC AAGAAGAAGA
901 AGCAGCGCAA GCACAGCGCC AGTGAGGAAG AGGCACGCGT GGGCGCAGGG
951 AAGAGGAAGC CGAGGCACCT GGAAGTTGAA ACAGATTCTA AGAAGAAAAA
1001 GATGAAGCTC CCAGAGCATC CTGAGGCGCG AGAACAGAA GACGATGAGG
1051 CTCTGCAAAA AGGTAAATTC AACTGGGAGG GAATATTAA AGCAATCTCT
1101 AACAGGGCCC CAGACAATGA AATAACCATC AAAAGCTTAA GGAAAAGGT
1151 TTTAGCTCAG TACTACACAG TGACAGATGA GCATCAGAA TCCGAAGAGG
1201 AACTCCTGGT CATCTTAAAC AAGAAAATCA GCAAGAACC TACCTTTAAG
1251 TTATTAAAGG ACAAGTCAA GCTTGTAAGA TGAACATTIG TGTATTTAAA
1301 AATTGAATCC ATTCTGCTGA CTTCTTCCTT TCACTGCTGT TTATAAAAATG
1351 TGTAATGAAT TCTAACAAC CAAATTTTGC TTTTGAAGC TGTATTTTAA
1401 AGTTAAGAAA ATATATTTT GGTATAACTT TTATGAGAAA AATAAAAATAT
1451 ATCTGGTCC AAACCTCAAA AAAAAAATAA AAAAAAATAA AAAAAAATAA
1501 AAA
```

## BLAST Results

-----

No BLAST result

## Medline entries

-----

93259460:

LYAR, a novel nucleolar protein with zinc finger DNA-binding motifs, is involved in cell growth regulation.

## Peptide information for frame 3

ORF from 144 bp to 1280 bp; peptide length: 379  
Category: strong similarity to known protein  
Classification: Cell division  
Prosites motifs: ATP\_GTP\_A (60-68)

```
1 MVFFTCNACG  ESKKIQVEK  HVSVCNCEC  LSCIDCGKDF  WGDDYKNHVK
51 CISEDQKYGG  KGYEGKTHKG  DIKQAWIQK  ISELIKRPNV  SPKVRELLEQ
101 ISAFDNVPRK  KAKFQNMWKN  SLKVHNESIL  DQVWNIFSEA  SNSEPVNKEQ
151 DQRPLHPVAN  PHAEISTKVP  ASKVKDAVEQ  QGEVKNKRE  RKEERQKRRK
201 REKKELKLEN  HQENSRNQKP  KRRKKGQEAD  LEAGGEEVPE  ANGSAQKRSK
251 KKKQKDSAS  EEARVGAGK  RRRRHSEVET  DSKKKMKLP  EHPEGGEPEP
301 DEAPAKGFN  WGTIKAILK  QAPDNEITIK  KLRKKVLAQY  YTVTDEHRS
351 EEELLVIFNK  KISKNPTFKL  LKDKVKLVK
```

## BLASTP hits

No BLASTP hits available

## Alert BLASTP hits for DKFZphtes3\_14g5, frame 3

PIR:A40683 cell growth regulating nucleolar protein LYAR - mouse, N = 1, Score = 1410, P = 2.7e-144

SWISSPROT:YQ58\_CAEEL HYPOTHETICAL 28.5 KD PROTEIN C16C10.8 IN CHROMOSOME III., N = 1, Score = 381, P = 2.9e-35

TREMBL:AC003058.18 gene: "F27F23.18"; product: "putative RNA-binding protein"; Arabidopsis thaliana chromosome II BAC F27F23 genomic sequence, complete sequence., N = 3, Score = 139, P = 4e-15

PIR:S70049 nucleic acid-binding protein YCR087c-a - yeast (Saccharomyces cerevisiae), N = 1, Score = 164, P = 1.4e-11

>PIR:A40683 cell growth regulating nucleolar protein LYAR - mouse  
Length = 388

## HSPs:

Score = 1410 (211.6 bits), Expect = 2.7e-144, P = 2.7e-144  
Identities = 275/388 (70%), Positives = 317/388 (81%)

```
Query:      1 MVFFTCNACGESVKKIQVEKHVSVCNCECLSCIDCGKDFWGDDYKNHVKCISEDQKYGG 60
            MVFFTCNACGESVKKIQVEK VS CRNCECLSCIDCGKDFWGDDYK+HVKCISE QKYGG
Sbjct:      1 MVFFTCNACGESVKKIQVEKQVSNCRNCECLSCIDCGKDFWGDDYKSHVVCISEGQKYGG 60

Query:     61 KGYEGKTHKGDIKQAWIQKISELIKRPNVSPKVRELLEQISAFDNVPRKKAKFQNMWKN 120
            KGYE KTHKGD KQAWIQKI+ELIK+PNVSPKVRELL+QISAFDNVP KAKFQNMWKN
Sbjct:     61 KGYEAKTHKGDQAWIQKINELIKRPNVSPKVRELLQISAFDNVPIKKAKFQNMWKN 120

Query:    121 SLKVHNESILDQVWNIFSEASNSEPVNKEQDQRPLHPVANPHAEIS-TKVPASKVKDAVE 179
            SLKVH++S+L+QVW+IFSEAS+SE ++Q Q P H A PHAE+ TKVP++K E
Sbjct:    121 SLKVHSDSVLEQVWDIFSEASSSE---QDQQPPSH-TAKPHAEMPITKVPSAKTNGTTE 176

Query:    180 QGEVKNKREKREERQKRRKREKKELKLENHQENSRNQKPKRRKKGQEADLEAGGEEVP 239
            +Q E KKKRERKEERQK RK+EKKEKLENHQEN R QKPKRRKK QEA EA GE+
Sbjct:    177 EQTEAKKKRERKEERQKRRKREKKELKLENHQENLRGQKPKRRKKQEAHEAAGEDGA 236

Query:    240 EANG-----SAGKRSKKKKQKDSASEEEA----RVGAGKRRK-RHSEVETDSKKKKM 287
            + +G G+ S++ R E+ A + AGKRRK +HS E+ KKKKM
Sbjct:    237 DSGSPPEKKKAQGGQASEEGADRNGGPGEDRAEGQTKTAAGKRRKPKHSGAESGYKKKKM 296

Query:    288 KLPEHPEGGEPEDEAPAKGFNWKGTIKAILKQAPDNEITIKLRKKVLAQYYTVTDEH 347
            KLPE PE GE +D EAP+KGKFNWKGTIKA+LKQAPDNEI++KKL+KKV+AQY+ V ++
Sbjct:    297 KLPEQPEEGEAKDHEAPSKGFNWKGTIKAVLKQAPDNEISVKKLKKKVIAQYHAVMNDT 356

Query:    348 HRSEEEELLVIFNKKISKNPTFKLLKDKVKLVK 379
            EEELL IFN+KIS+NPTFK+LKD+VKL+K
Sbjct:    357 SHHEEELLAIFNRKISRNPFTFKVLKDRVKLLK 388
```

## Pedant information for DKFZphtes3\_14g5, frame 3

```
[LENGTH]      379
[MW]           43634.03
[PI]           9.59
[HOMOL]        PIR:A40683 cell growth regulating nucleolar protein LYAR - mouse 1e-122
[FUNCAT]       04.99 other transcription activities [S. cerevisiae, YCR087c-a] 2e-11
[BLOCKS]       BL00603D Thymidine kinase cellular-type proteins
[BLOCKS]       BL00530C
[PROSITE]      ATP_GTP_A      1
[KW]           All_Alpha
[KW]           LOW_COMPLEXITY  18.73 %
```

```
SEQ      MVFFTCNACGESVKKIQVEKHVSVCRNCECLSIDCGKDFWGDDYKHNHVKCISEDQRYGG
SEG
PRD      cccccccccccccchhhhhhhheeeccccceeeccccccccccccccccceeeccccccccc

SEQ      KGYEGKTHKGDIKQQAOWIKISELIRKPNVSPKVRLELEQISAFDNVPRKKAKFQNMWKN
SEG
PRD      cccccccccchhhhhhhhhhhhhhhhhhhhhcchhhhhhhhhhhhhccccchhhhhhhhhhhhhc

SEQ      SLKVHNESILDQVMNIFSEASNSEPVNKEQDQRPLHPVANPHAETSKVPASKVKDAVEQ
SEG
PRD      cccccchhhhhhhhhhhhhhhhhhhhhcchhhhhhhhhhhccccccccccccceeeccccchhhhh

SEQ      QGEVKKNKRRERKEERQKKRREKKELKLENHQENSRNQPKPKRKKGQEADLEAGGEEVPE
SEG
PRD      . . . . . xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx . . . . . xxxxxxxxxxxxxxx
      hhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhccccccccccchhhhhhhhhhhcccccccc

SEQ      ANGSAKGRSKKKQRKDSASEEARVAGAKRRRRHSEVETDSKKKKMKLPEHPGEGPED
SEG
PRD      . . . . . xxxxxxxxxxxxxxxxxxxxxxx . . . . . xxxxxxxxxxxxxxx
      cccccccccchhhhhhhhhcchhhhhhhhhhhccccccccccccccchhhhhhhcccccccccc

SEQ      DEAPAKGKFNWGTIKAILKQAPDNEITIKKLKKVLQAYTYTDEHHRSEELLVIFNK
SEG
PRD      xxxxxx
      cccccceeehhhhhhhhhhhhccccccccchhhhhhhhhhhhhhhhhhhcchhhhhhhhhhhhhhh

SEQ      KISKNPFTKLLDKVKVLVK
SEG      . . . . . xxxxxxxxxxxxxxx
PRD      cccccccchhhhhhhhhhhccc
```

PS00017      60->68    ATP GTP A      PDOC00017

(No Pfam data available for DKFZphtes3 14q5.3)

DKFZphtes3\_14h21

-----

group: nucleic acid management

DKFZphtes3\_14h21 encodes a novel 648 amino acid protein with strong similarity to mus musculus RNA helicase and several RNA-dependent ATPases from the DEAD box family.

RNA helicases comprise a large family of proteins that are involved in basic biological systems such as nuclear and mitochondrial splicing processes, RNA editing, rRNA processing, translation initiation, nuclear mRNA export, and mRNA degradation. RNA helicases are essential factors in cell development and differentiation, and some of them play a role in transcription and replication of viral single-stranded RNA genomes. The members of the largest subgroup, the DEAD and DEAH box proteins, exhibit a strong dependence of the unwinding activity on ATP hydrolysis. The novel protein contains a DEAD-box and a ATP/GTP-binding site motif A (P-loop) and is a new member of this subgroup.

The new protein can find application in modulating RNA metabolism and gene expression.

strong similarity to RNA helicases

start at Bp 33 matches Kozak consensus ACNATg

Sequenced by BMFZ

Locus: unknown

Insert length: 2200 bp

Poly A stretch at pos. 2166, polyadenylation signal at pos. 2140

```
1 CAACGACGTC GGACGCGCCC CTTCTTGGA CAATGTCCCA CCACGGAGGA
51 GCTCCCAAGG CCTCTACGTG GGTGCTGCT AGTCGGCGAA GCTCCGACGT
101 GTCCCGAGCG CCAGAGAGGA GGCCGGCGGA GGAGTTGAAT CGAACAGGTC
151 CTGAGGGATA TAGTGTCCGC AGAGGTGTC GCTGGAGAGG CACCTCTAGG
201 CCCCAGGAGG CCGTGGCCGC TGGTCAGGAG GAACGCGCGC TGTGTTTGGC
251 TTTGAAGAGC CACTTTGTTG GCGCGGTAAT CGGTGCTGGT GGGTCAAAAA
301 TAAAGAATAT ACAAAGTACA ACAAACACCA CAATCCAAAT AATACAAGAA
351 CAACCAAGAT CATTAGTCAA AATTTTGGC AGCAAGGCAA TGCAACGAA
401 AGCAAAAGCA GTGATAGACA ATTTTGTAA AAAGCTAGAA GAAAATTACA
451 ATTCAGATG CGGAATTGAT ACTGCATTCC AACCTTCTGT TGGAAAGAT
501 GGAAGCAGAG ATAACAATGT TGTGCGAGGA GATCGGCCAT TGATAGATTG
551 GGATCAAAAT AGAGAGGAAG GTTTGAAATG GCAAAAAACA AAGTGGGCAG
601 ATTTACCACC AATTAAGAAA AACTTTATA AAGAGTCCAC TGCCACAAGT
651 GCCATGTCAA AAGTAGAAGC AGATAGTTGG AGGAAGAGAA ATTTTAATAT
701 AACGTGGGAT GACTTGAGAG ATGGGGAGAA ACGACCTATC CCCAATCCTA
751 CCTGCACATT TGATGACGCC TTTCAATGTT ATCCTGAGGT TATGGAAAAC
801 ATTAAGAGG CAGGTTTCCA AAAGCCAACA CCTATTCAGT CACAGGCATG
851 GCCCATTGTG TTGCAAGGAA TAGATCTTAT AGGAGTAGCC CAGACTGGAA
901 CAGGAAAGAC ATTGTGTTAT TTAATGCCTG GATTATTCA TCTGGTCCTT
951 CAACCCAGCC TTAAGGTCA AAGGAATAGA CCCGGCATGT TAGTTCTAAC
1001 TCCCACTCGG GAATTAGCAC TTCAAGTAGA AGGAGAATGT TGCAAAATAT
1051 CATATAAAGG GCTTCGGAGT GTTTGTGAT ATGGTGGTGG AAATAGAGAT
1101 GAACAAATAG AAGAGCTTAA AAAAGGTGTA GATATCATAA TTGCAACTCC
1151 CGGAAGATTG AATGATCTGC AAATGAGTAA CTTGCTCAAT CTGAAGAATA
1201 TAACCTACTT GGTTTTAGAT GAAGCAGACA AGATGTTGGA CATGGGATTT
1251 GAACCCAGCA TAATGAAGAT TTTGTTAGAT GTGCGCCAG ATAGGCAGAC
1301 AGTTATGACC AGTGCTACAT GGCCCTATTC AGTTCATCGC CTCGCACAAT
1351 CTTATTTGAA AGAACCAATG ATTGCTATG TTGGTACATT GGATCTAGTT
1401 GCTGTAAATT CAGTGAAGCA AAATATAATT GTAACCAACG AGGAAGAGAA
1451 ATGGAGTCAC ATGCAAACTT TTCTACAGAG TATGTCATCC ACAGACAAAG
1501 TCATTGTCTT CGTTTCTCGA AAAGCTGTG CCGATCACTT ATCAAGTGAC
1551 CTAATACTTG GAAATATATC AGTAGAGTCT CTGCATGGAG ATAGAGAACA
1601 GAGAGATCGG GAGAAAGCAT TAGAGAACTT TAAACAGGCG AAAGTGAGAA
1651 TACTAATTGC AACTGATCTA GCCTCTAGAG GACTTGATGT CCATGACGTT
1701 ACACATGTCT ATAATTTTGA CTTTCCACGG AATATTGAAG AATACGTACA
1751 CCGAATAGGG CGCACGGGAA GAGCAGGGAG GACTGGTCTT TCCATTACAA
1801 CTTTGACTAG AAATGATTGG AGGGTTGCCT CTGAATTGAT TAATATTCTG
1851 GAAAGAGCAA ATCAGAGTAT TCCAGAGGAG CTTGTATCAA TGGCTGAGAG
1901 GTTTGAGGCA CATCAACGGA AAAGGGAAT GGAAGAGAAA ATGGAAAGAC
1951 CTCAGGAAG GCCCAAGAA TTTCAATTAAT GTCTTCTGTA CTAGTGGGGT
2001 AGAGAATTCA AGATTTTTTA GAAATATAGT AAGACAGAAG TATTGGACAT
2051 GTTGGCAGTA TGAAGAGACC GGACTGATTT GACTGATTCT TAAATAATA
2101 GTGTTTGAAA ATATAGAATC CAGTGTTTAA TACTTCTTTT AATAAAAAATA
2151 GAAGTATTTA AACTTGAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA
```

BLAST Results

-----

## No BLAST result

## Medline entries

No Medline entry

## Peptide information for frame 3

ORF from 33 bp to 1976 bp; peptide length: 648  
 Category: strong similarity to known protein  
 Classification: Nucleic acid management  
 Prosite motifs: ATP\_GTP\_A (286-294)  
 DEAD\_ATP\_HELICASE (394-403)

```

1 MSHHGGAPKA STWVVASRRS STVSRAPERR PAEELNRTGP EGYSVGRGGR
51 WRGTSRPPEA VAAGHEELPL CFALKSHFVG AVIGRGGSKI KNIQSTNTT
101 IQIIQEPES LVKIFGSKAM QTKAKAVIDN FVKKLEENYN SECIGDTAFQ
151 PSVKGKGSST NNVVAGDRPL IDWDQIREEG LKWKQTKWAD LPPIKKNFYK
201 ESTATSAMSK VEADSWRKEN FNITWDDLKD GEKRPINPT CTFDDAFQCY
251 PEVMENIKKA GFQKPTPIQS QAWPIVLQGI DLIGVAQTCT GKTLCYLMGP
301 FIHLVLQPSL KGQRNRPGLM VLTPTRELAL QVEGECCCKYS YKGLRSVCVY
351 GGGNRDEQIE ELKKGVDI II ATPGRLNDLQ MSNFVNLKNI TYLVLDEADK
401 MLDMGFEPQI MKILLDVRPD RQTVMTSATW PHSVHRLAQS YLKEPMIVYV
451 GTLDLVAVSS VKQNIIVTTE EEKWSHMQTF LQSMSSTDKV IVFVSRKAVA
501 DHLSSDLILG NISVESLHGD REQDRREKAL ENFKTGKVR I LIAIDLASRG
551 LDVHDVTHYV NFDFFRNIEE YVHRIGRTGR AGRTGVSITT LTRNDWRVAS
601 ELINILERAN QSIPEELVSM AERFEAHQRK REMERKMERP QGRPKKFKH

```

## BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphtes3\_14h21, frame 3

TREMBL:CEY54G11A\_9 gene: "Y54G11A.3"; Caenorhabditis elegans cosmid  
 Y54G11A, N = 1, Score = 1008, P = 1.1e-101

TREMBL:SPBP8B7\_16 gene: "dbp2"; "SPBP8B7.16c"; product: "p68-like  
 protein."; S.pombe chromosome II pl p8B7., N = 1, Score = 971, P =  
 9.1e-98

PIR:S13757 RNA helicase DBP2 - yeast (Saccharomyces cerevisiae), N = 1,  
 Score = 970, P = 1.2e-97

PIR:S14048 RNA helicase dbp2 - fission yeast (Schizosaccharomyces  
 pombe), N = 1, Score = 961, P = 1e-96

PIR:A57514 RNA helicase HEL117 - rat, N = 2, Score = 888, P = 7.8e-91

>TREMBL:CEY54G11A\_9 gene: "Y54G11A.3"; Caenorhabditis elegans cosmid  
 Y54G11A  
 Length = 504

## HSPs:

Score = 1008 (151.2 bits), Expect = 1.1e-101, P = 1.1e-101  
 Identities = 211/473 (44%), Positives = 298/473 (63%)

```

Query: 174 DQIREGLKWQTKWADLPPIKKNFYKESTATSAMSKVEADSWRKENFNITWDDLKDGEK 233
      D+++E W K PI ++ YK +S + + ++
Sbjct: 23 DRLKDNFSWMK-----PIVRDLYKIPNEQKNLSPEQLQELYTNGGVMKVYPFREEST 75

Query: 234 RPIPNPTCTFDDAFQCYPEVMENIKKAGFQKPTPIQSQAWPIVLQGI DLIGVAQTGTGKT 293
      IP P +F+ AF +M I+K GF+KP+PIQSQ WP++L G D IGV+QTG+GKT
Sbjct: 76 VKIPPPVNSFEQAFGSNASIMGEIRKNGFEKPSPIQSQMWPLLSSGQDCIGVSQTGSGKT 135

Query: 294 LCYLMPGFIHLVLQPSL-----KGQRNRPGLMVLTPTRELAQVEGECCCKYSYKGLRSVC 348
      L +L+P +H+ Q + + Q+ P ,LVL+PTRELA Q+EGE KYSY G +SVC
Sbjct: 136 LAFLLPALLHIDAQLAQYEKNDEEQKPSFVLVLSPTRELAQQIEGEVKKYSYNGYKSV 195

Query: 349 VYGGNRDEQIEELKKGVDI IIATPGRLNDLQMSNFVNLKNITYLVLDEADKMLDMGFEP 408

```

```

          +YGGG+R EQ+E  + GV+I+IATPGRL DL      ++L ++TY+VLDEAD+MLDMGFE
Sbjct:  196 LYGGGSRPEQEACRGGEIATPGRLTDLSDGVLASVTVVVLDEADRLDMGFEV 255

Query:  409 QIMKILLDVRPDRQTVMTSATWPHSVHRLAQSYLKEPMIVYVGTLDLVAVSSVKQNIIVT 468
          I ++IL ++RPDR  +TSATWP V +L  Y KE ++  G+LDL +  SV Q
Sbjct:  256 AIRRILFEIRPDRLVALTSATWPEGVRKLTDKYTKAAMVAVNGSLDLSCKSVTQFFFEV 315

Query:  469 TEEKW---SHMQTFLOQMSSTD-KVIVFVSRKAVADHLSSDLILGNISVESLHGDREQR 524
          + ++      + FL +      + K+I+V K +ADHLSSD + I+ + LHG R Q
Sbjct:  316 PHDSRFLRVCEIVNFLTAAHGQNYKMIIFVKSVMADHLSSDFCMKGINSQGLHGGRSQS 375

Query:  525 DREKALENFKTGKVRILIATDLASRGDLVDVTHVYNDFPRNIEEYVHRIGRTGRAGR 584
          DRE +L  ++G+V+IL+ATDLASRG+DV D+THV N+DFP +IEEYVHR+GRTGRAGR
Sbjct:  376 DREMSLNLRSGEVQILVATDLASRGIDVPDITHVLNDFPMDIEEYVHRVGRTRGRGRK 435

Query:  585 GVSITTLTRNDWRVASELINILERANQSIPEELVSMAERFEAHQRKREMERKMERPOGRP 644
          G +++ L ND  LI ILE++ Q +P++L AE++      K + R  RP R
Sbjct:  436 GEAMSFLLWNDRSNFEGLIQILEKSEQEVPDQLRRDAEKYRL---KQSGRDGPRPFRN 492

Query:  645 KK 646
          K
Sbjct:  493 NK 494

```

Pedant information for DKFZphtes3\_14h21, frame 3

#### Report for DKFZphtes3\_14h21.3

```

[LENGTH]      648
[MW]           72873.51
[pI]           8.84
[HOMOL]        TREMBL:CEY54G11A_9 gene: "Y54G11A.3"; Caenorhabditis elegans cosmid Y54G11A 1e-101
[FUNCAT]       04.01.04 rna processing [S. cerevisiae, YNL112w] 2e-97
[FUNCAT]       30.10 nuclear organization [S. cerevisiae, YNL112w] 2e-97
[FUNCAT]       04.05.03 rna processing (splicing) [S. cerevisiae, YPL119c] 4e-72
[FUNCAT]       30.03 organization of cytoplasm [S. cerevisiae, YOR204w] 2e-70
[FUNCAT]       05.04 translation (initiation, elongation and termination) [S. cerevisiae, YOR204w] 2e-70
[FUNCAT]       06.10 assembly of protein complexes [S. cerevisiae, YBR237w] 1e-61
[FUNCAT]       1 genome replication, transcription, recombination and repair [H. influenzae, HI0892] 2e-49
[FUNCAT]       j rna translation and ribosome biogenesis [H. influenzae, HI0231 RNA] 1e-48
[FUNCAT]       04.99 other transcription activities [S. cerevisiae, YDL160c] 9e-45
[FUNCAT]       04.05.01.07 chromatin modification [S. cerevisiae, YMR290c] 3e-44
[FUNCAT]       09.01 biogenesis of cell wall [S. cerevisiae, YJL033w] 2e-36
[FUNCAT]       98 classification not yet clear-cut [S. cerevisiae, YOR046c] 7e-32
[FUNCAT]       30.16 mitochondrial organization [S. cerevisiae, YDR194c] 2e-28
[FUNCAT]       99 unclassified proteins [S. cerevisiae, YGL064c] 5e-10
[FUNCAT]       11.10 cell death [S. cerevisiae, YMR190c] 2e-08
[FUNCAT]       03.19 recombination and dna repair [S. cerevisiae, YMR190c] 2e-08
[FUNCAT]       r general function prediction [M. jannaschii, MJ1401] 1e-07
[BLOCKS]       BL00039D DEAD-box subfamily ATP-dependent helicases proteins
[BLOCKS]       BL00039C DEAD-box subfamily ATP-dependent helicases proteins
[BLOCKS]       BL00039B DEAD-box subfamily ATP-dependent helicases proteins
[BLOCKS]       BL00039A DEAD-box subfamily ATP-dependent helicases proteins
[PIRKW]        nucleus 4e-96
[PIRKW]        RNA binding 3e-87
[PIRKW]        DEAD box 5e-50
[PIRKW]        transmembrane protein 4e-27
[PIRKW]        DNA binding 3e-67
[PIRKW]        recF recombination pathway 3e-10
[PIRKW]        ATP 4e-96
[PIRKW]        purine nucleotide binding 5e-50
[PIRKW]        P-loop 4e-96
[PIRKW]        hydrolase 9e-45
[PIRKW]        protein biosynthesis 5e-50
[PIRKW]        ATP binding 1e-61
[SUPFAM]       WW repeat homology 8e-88
[SUPFAM]       DEAD/H box helicase homology 4e-96
[SUPFAM]       unassigned DEAD/H box helicases 7e-87
[SUPFAM]       ATP-dependent RNA helicase DBP1 4e-96
[SUPFAM]       ATP-dependent RNA helicase DHH1 2e-43
[SUPFAM]       recQ protein 3e-10
[SUPFAM]       Bloom's syndrome helicase 5e-07
[SUPFAM]       translation initiation factor eIF-4A 5e-50
[SUPFAM]       recQ helicase homology 3e-10
[SUPFAM]       tobacco ATP-dependent RNA helicase DB10 8e-88
[PROSITE]      DEAD_ATP_HELICASE 1

```

```

[PROSITE]      ATP_GTP_A      1
[PFAM]         Helicases conserved C-terminal domain
[PFAM]         KH domain family of RNA binding proteins
[PFAM]         DEAD and DEAH box helicases
[KW]           Alpha_Beta
[KW]           LOW_COMPLEXITY      8.49 %

SEQ      MSHHGGAPKASTWVVASRRSSTVSRAPERPAEELNRTGPEGYSVGRGGRWGTSRPPEA
SEG      .....XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
PRD      cccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc

SEQ      VAAGHEELPLCFALKSHFVGAVIGRGGSKIKNISTNTTIQIIQEPESLVKIFGSKAM
SEG      .....XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
PRD      cccccccccchhhhhccccccccccccccccccccccccccccccccccccccccchh

SEQ      QTKAKAVIDNFVKLEENYNSECIDTAFQPSVGKDGSTDNNVAGDRPLIDWDQIREEG
SEG      .....
PRD      hhhhhhhhhhhhhhhhhhhhhccccccccccccccccccccccccccccccccccccccc

SEQ      LKWQKTWADLPPIKKNFYKESTATSAMSKEADSWRKENFNITWDDLKDGKRPIPNPT
SEG      .....
PRD      chhhhhhhccccccccccccccccccccchhhhhhhhhhhhhhhheeeccccccccccccccc

SEQ      CTDDAFQCYPEVMENIKKAGFQKPTPIQSQAWPIVLQIDLIGVAQTGTGKTLCYLMPG
SEG      .....
PRD      cccccccccchhhhhhhhhhhhhccccccccccccccccccccccccccccccccccccccc

SEQ      FIHLVLQPSLKGQRNRPGLVLTPTRELALQVEGECCKYSYKGLRSVCVYGGGNRDEQIE
SEG      .....
PRD      eeeccccccccccccccccccccchhhhhhhhhhhhhhhccccccccccccccccchhh

SEQ      ELKKGVDIITATPGRLNDLQMSNFVNLKNITYLVLEADKMLDMGFEPQIMKILLDVRPD
SEG      .....
PRD      hhhheeeccccccccchhhhhhhccccccccccccchhhhhhhhhhhhhccccccccchhh

SEQ      RQVTMTSATWPHSVHRLAQSYLKEPMIVYVGTLDLVAVSSVKQNIIVTTEEEKWSHMQTF
SEG      .....
PRD      ceeeeccccchhhhhhhhhhhhhheeeccccccccccccccccccccchhhhhhhhhhh

SEQ      LQSMSTDKVIVFVSRKAVADHLSSDLILGNISVESLHGDREQRDRKALENFKTGKVRI
SEG      .....
PRD      hhhccccccccccccchhhhhhhhhhhhhccccccccccccccccchhhhhhhhhhhccccce

SEQ      LIATDLASRGLDVHDVTHVYNFDFPRNIEEYVHRIGRTGRAGTGVSTITLNRNDWRVAS
SEG      .....XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
PRD      eeehhhhhhccccccccccccccccccccccccccccccccccccccccccccccccchhh

SEQ      ELINILERANQSIPEELVSMAERFEAHQRKREMERKMERPQGRPKKFH
SEG      .....XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
PRD      hhhhhhhhhccccccccchhhhhhhhhhhhhhhhhhhhhhhccccccccccccccccccc

```

## Prosites for DKFZphtes3\_14h21.3

PS00017	286->294	ATP_GTP_A	PDOC00017
PS00039	394->403	DEAD_ATP_HELICASE	PDOC00039

## Pfam for DKFZphtes3\_14h21.3

HMM_NAME	DEAD and DEAH box helicases	
HMM	*gLpPWILRniyeMGFEkPTPIQQaIPiLeGRDVMACAQTGSGKTAAP P++++NI+++GF KPTPIQ+QA+PI+L+G D+++ AQTG+GKT+++	
Query	248	QCYPEVMENIKKAGFQKPTPIQSQAWPIVLQIDLIGVAQTGTGKTLCY 296
HMM	LIPMLQHIDwdfWpqpQd..PrALILAPTRELAMQIEEcRkFgkHmNg L+P ++H+ +P +++ Q+ P +L+L+PTRELA+Q++ EC K+++ +	
Query	297	LMPGFIHLVLQP-SLKGQRNRPGLVLTPTRELALQVEGECCKYSYK-G- 343
HMM	IRImcIYGGtnMRdQMRmLeRGpPHIVIATPGRLIDHIERgtldLDRIeM +R++C+YGG N ++Q+++L++G+ +I+IATPGRL D+ +++ ++L++I++	
Query	344	LRSVCVYGGGNRDEQIEELKGV-DIIATPGRLNDLQMSNFVNLKNITY 392
HMM	LVMDEADRMMLDMGFIDQIRrIMrQIPmwnRQTMFSATMPdeIqELARr LV+DEAD+MLDMGF++QI++I+ ++ ++RQT+M SAT+P ++ +LA	
Query	393	LVLDEADKMLDMGFEPQIMKILLDVR--PDRQVTMTSATWPHSVHRLAQS 440



```

HMM          FMRNPIRInId.MdElTtnEnIkQwYiyVerEMWKfdClerLie*
++++P  + ++ D  +++ +KQ +I+  E++K +  ++++
Query      441 YLKEPMIVYVGTLDLVAVS-SVKQNIIVTT-EEKWSHMOTFLQ  482

HMM_NAME     KH domain family of RNA binding proteins

HMM          *rIiIPedhMGMIIGKGGsNIRqIREEYgvrINIPdecCeDstdRIITit
+ + ++++G++IG+GGS I++I++ ++++I I++E+ +  + I
Query      71  CFALKSHFVGAVIGRGGSKIKNIQSTTNTTIQIIQEQ-P---ESLVKIF  115

HMM          G*
              G
Query      116 G  116

HMM_NAME     Helicases conserved C-terminal domain

HMM          *EileeWlkn1...GlrvmYIHGdMpQeERdeIMddfNnGEynVLicTD
+ +++ L+ +  +I+V ++HGD++Q++R+++++F++G+ ++LI+TD
Query      497 KAVADHLSSDLILGNISVESLHGDRQDRREKALENFKTGKVRILIATD  545

HMM          VggRGIDIPdVNHVINYDMPWNPEqYIQRIGRTgRIG*
+++RG+D+ DV HV+N+D+P+N+E Y++RIGRTGR+G
Query      546 LASRGLDVHVDVTHVYNFDFPRNIEEYVHRIGRTGRAG  582

```

DKFZphtes3\_14p14

group: testes derived

DKFZphtes3\_14p14 encodes a novel 159 amino acid protein without similarity to known proteins.

No informative BLAST results; No predictive prosite, pfam or SCOP motive.

The new protein can find application in studying the expression profile of testis-specific genes.

unknown

complete cDNA, complete cds, few EST hits

Sequenced by BMFZ

Locus: unknown

Insert length: 3969 bp

Poly A stretch at pos. 3948, polyadenylation signal at pos. 3927

```
1 GAAGCCCAGG CTCTCCTTAG TTGACTGTGT GTTAATCACC CAGCAATTTT
51 ATTACTCAAC AGCTCTCCAG AGTTGCACAT TACAGCTGGG GTAGAAATTT
101 GGTGCTGAAG GCCAGGCAGA GCATTGGCT GTAGGGAGGC CGATCCTCCT
151 CGGCGCTGTT ACCGGCGGCT CTTTGTCTT AGACCTGGGG TTCTGGGCTT
201 CACGGATTCC AAGGAATGGA ACCTTGGGCC ATGCGTGTGA ACGAGCTCTA
251 TGTGATGAC CCAGACAAGG ACAGCGGTGG CAAGATCGAC GTCAGTCTGA
301 ACATCAGTTT ACCCAATCTG CACTGCCGAGT TGGTTGGGCT TGACATTTCAG
351 GATGAGATGG GCAGGGCAGA AGTGGGCCAC ATCGACAAC TCCATGAAGAT
401 CCCGCTGAAC AATGGGGCAG GCTGCCGCTT CGAGGGGGCAG TTCAGCATCA
451 ACAAGGTATG GAAGCCCTGC CTCAGCCCTT TCTACCTGCT CCCCTTTTCT
501 GCTGTCTCCC CGCTCCCTGG AAACCTGGTTG TGGAGGCACT CACTCGACCT
551 GACCCCTGACA CAGCCCCCAG CAAGCGAGGG TTGCTGTCCA GCTGCCTTGG
601 CGTTCCTGCT GAGAATCTGG ATGGGGGTCC AGGCTCCCTG GGGTTTTAAG
651 CCCCTGATGG CTGGTTCAGG AAGGAGCTAC TCTTCTCTCC AGTGAGGGGG
701 ACAATGATGA GAAGACCTGA GGATTTCAG CCCCAGCCC TGGGTTCAAG
751 TCCAGGCTCT ACCCCTTCTT GGGCCCTACA AGTCACTTGA CCCATCTTAG
801 GCTGAGGGTG TGATGGCGAT AATAGTATCA CGATACCACC CACTTCACAA
851 AGTTTGTGTG GGGATTAAT GAGCTAATGC AGATTTCATT ATTCAGAAAA
901 ATTTTGAAT GGCACGTTCT GTGTTCCAGG GTCGGTGATA GGCTCTGGGG
951 CAGCGTCTCT GGGCTGGTGG GGCTCCCAT TCGGTAGAGG GAGACAGTCT
1001 ACAAAACCAGA AAGCATCAGG GATGCTAAGT GCAGTGATGA GGAATAAAGC
1051 CAAGGGGAGT GAGATGAGT GGGCTTGAAG GTACCTTGTC CGCTCAGAAG
1101 GACCAATCAA GGTTCCTGT TGTTTGTCC TCAGAACCAAG GAGCTTCAGA
1151 TCCTAAGTCA AGTGGGTGAA CGCAGTGCCC TTGGGAGGGC CGAGGCACCC
1201 GGTGGCAGCT GGCAGGGTTT TGCTCAGCAC GTGCCGGCCT TCCTCGAAGC
1251 TCGGTACTGT CACAGTGGAG CCTCTCAACA ACCTGTGTAG GCAGCACCAT
1301 TTGACAGGTT AGGATGCTGG GGGCCAGAGA GGTAAAGTGT CTTGCCCGAG
1351 GTCACACAGC TATCTGCATG TCCCACAAC TCCCTTCCCA GCCCCAGCCA
1401 AACTGAGCCA CTGGCCACTC CTGGCTTCTC CTTGTCCCTC CTGCAGCCTC
1451 TGCTCAGAAC GCCCTTCTCT CAGACCTTGA CACTGTAGCT GGGGTTGCAA
1501 AGTCACTGGC CACATCCAGC CCAAAGATAA ATTTTGTGTT TCCAGTATAG
1551 CATTTAACTG CATCAGAAC AGTATGAAAA GACCAGGAAT CCAGATTTCCT
1601 GGCTTTTAAA AGTCAGAGGC TCTCACTACA CTGGGTCCGT GTTCCCGCTA
1651 TGACAATGAC CTGGCACCAG TGGGCAGTGT TCCCTTTAG AGAGGGTGTG
1701 TGCTGTCCCT TCCCACAGTC CCTGGCAGGC GGCTGGAAGG CCAGGCCCTGG
1751 TCATCTGTCA AGCAGGGTGG ACTTCTTACG TGACAGTTCA GGGCTCCCTT
1801 AAGTGTCTAA GCAGAAGCTG CAAGGCTTTC TTAAGGTTTC GAGTGTGTCT
1851 GGGGAAATCT TGCTGCATGT TGTGGGTTAA AGGGAGTCTC TCACCAGCCC
1901 AGGCCCTCAG GAGGAGGAGA TACCAGGAGG CAGGGATGCT GGGGGTCTGT
1951 GTTCACTGGG GGCTCTCTCT GCCATGAGC TGCCACACAG CACCTTTGCC
2001 ATGCCCCGTA ATTTGGATT TATGGTGGTT GTGATGAAA GCCATTGAG
2051 GGTTTTGAAC AGGGAGGCAA TGTAACTAGA TTTATGCTT AGAATGGAC
2101 TATCCCATAG GTTGCCACCA GCCACATAAG GCTATTTAAA TTAATTCAAA
2151 TTAATGTAC AATTCACTCA CTCATTCTCA TCAACCACAT TTCAGTGTCT
2201 CAAAGCCAGC TGCTGGCTAG GGGCCACAGC GTTAGACAGT GCAGAGAGAA
2251 AGCACTTCCA TCGCTGAGGA AAGTTCTGCT GGACCGCACA CCCTTAGAAG
2301 GATGGCTCTG GTGGCCGGGC GCGGTGGCTC AAACCTGTAA TCCCAGCACT
2351 TTGGGAGGCC GAGGTGGGTG GATCACGAGG TCAGGAGATC GAGACCATCC
2401 CGGCTAACAT GGTGAAACCC TGCTCTACT AAAAAACAA AAAAAACAA
2451 AATTAGCCGG GCGTGGTTGC GGGCACCTGT AGTCCAGCT ACTCAGGAGG
2501 CTGAGGCGGG AGAATGGCAT GAACCCGGGA GGTGGAGCTT GCAGTGAGCC
2551 AAGATCTGAT CACTGCACTC CAGTCTGGGC GACAGAGTGA GACTCATCT
2601 CAAACAAAC AAAAAAGGA TGGGGCTGGG CTGGAGAGGG TGGCAGGCAG
2651 TGGTTGTGGC AGTGGAGCTG GGGAGATGTG GTCGGATTAG GGAGGTAGAA
2701 TCAATAAGAC TCAGTGAAGA ATCGGATGTG GGGGTAAGG CACATGTGGA
```

2751 AGCAAAGAAA CCTTTGACGT CTTTGTCTTG ACAACCGGGT GGTCTGTGTT  
 2801 CTAGACATGG AAGCTTAGAA AAGCCTGGAG TCTGTGGGAA GTAGGTAGGG  
 2851 CTGGGCACTG GTCATTCCAC TCTGGTTTCC TTTGGGGTTC CCATTAGGTG  
 2901 TCTACAGGGA GAGGTGAAAT TGGAAAGTGG AGGTGTGGAG AGTTCAGGAG  
 2951 AGGGTCTCTG ACCACAGATG TTGAGGTGGG AGTCATTAGT GAATAGATGA  
 3001 TGTGGAAAGT CATGGGTCCT CAGAGTGGGG GCTCCTTAAG CCTCCAGGCC  
 3051 AGCAGCATCA GCATCACCTG GGAGATTGTT AGGAATGCAG ATTCTCAGGC  
 3101 CCCCCTAAGA CCCACCGACT CTGTGCTAGA ACAAGCGCCC CTCAGAGATT  
 3151 CTGATGCCAC TGAAGTTTGA GGAGCATTGG TTTAAGCAAG ATTACCTACG  
 3201 GAGAGGCTGT AGATCCGTGT TCTAAACCTG GGGTCCACAG ACACCCCAA  
 3251 GAAGAGCGGA TTGAATGCAA GAGATCTATG AAGTTGGATG GGGGAAAAAT  
 3301 TGACATCTTT ATTTTGTGTA AACTCGATCT AAAGTTTAGC ATTTCCATCT  
 3351 GCGATGAATG TAGGCCACAA ACCACAGTAG TATTAGCAGT GCCTGGGACC  
 3401 TCCTCAACAA CAGAAATTGC CGGTATTAT AGCACGTTAC AGTTGTTGCA  
 3451 GATAATTTCC AGAGACTGTT TATATGCACC ACTGTTTAA AATTACGGTG  
 3501 ATTGGCCAGG TGCAGTGGCT CACACCTGTA ATCCAGCAC TTTGGGAGGC  
 3551 CAAAGTGGGT GGATCACTTG AGGAGTTCAA GACCAGCCTG GTCAACATGT  
 3601 CAAAACCCCTG TATCTACAAA AAAATACAAA AGTTAACCAG GCCTATGCTT  
 3651 GTAGTCACAG CTACTCGGGA GGCCGAGGTG GGAGGGCTTT CTGAGCCAG  
 3701 GGAGGTAGAG GCTTCAGTGA GCTGAGATCG CACCACCACA CTCCAGCCTG  
 3751 GGTGACAGAG TGAACCCCTT AATCAATCAG TCAATAAAAA TTACAGTAAT  
 3801 TATTAGACCC ACCACTAGGT CATCTTATTT GATGCATCAG TAAAGCAGCA  
 3851 TATTCAAATG TGGATTTTAA AATATTTTAA TTACTATTTA AATATCTCTT  
 3901 TACTTTGTAA TCCTATGCAT TTTACGCATT AAAACATTTT AAGCATTTAA  
 3951 AAAAAAAAAA AAAAAAAAAA

BLAST Results

No BLAST result

Medline entries

No Medline entry

Peptide information for frame 3

ORF from 216 bp to 692 bp; peptide length: 159  
 Category: putative protein  
 Classification: no clue

1 MERWAMRVNE LYVDDPKDSD GSKIDVSLNI SLPNLHCELV GLDIQDEMGR  
 51 HEVGHIDNSM KIPLNNGAGC RFEGQFSINK VWKPLSPFY LLPFAVSPL  
 101 PGNWLWRHSL DLTLTQPPAS EGSCPAAWPF LLRIWMGVQA PWGFKPLMAG  
 151 SGRSYSSLQ

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphtes3\_14p14, frame 3

No Alert BLASTP hits found

Pendant information for DKFZphtes3\_14p14, frame 3

Report for DKFZphtes3\_14p14.3

{LENGTH} 159  
 {MW} 17778.55  
 {pI} 5.74  
 {FUNCAT} 99 unclassified proteins [S. cerevisiae, YAL042w] 5e-04  
 {KW} Alpha\_Beta

SEQ MERWAMRVNELYVDDPKDSDGSKIDVSLNISLPNLHCELVGLDIQDEMGRHEVGHIDNSM  
 PRD ccchhhhhhhcc

SEQ KIPLNNGAGCRFEGQFSINKVWKPLSPFYLLPFAVSPLPGNWLWRHSLDLTLTQPPAS  
 PRD eeecc

WO 01/12659

PCT/IB00/01496

SEQ EGSCPAAWPFLRLRIWMGVQAPWGFKPLMAGSGRSYSSLQ  
PRD cccccchhhhhhhhhhhcccccccccccccccccccc

(No Prosite data available for DKFZphtes3\_14p14.3)

(No Pfam data available for DKFZphtes3\_14p14.3)

DKFZphtes3\_14p7  
-----

group: testes derived

DKFZphtes3\_14p7 encodes a novel 702 amino acid protein with very weak similarity to kinesin associated protein KAP3.

No informative BLAST results; No predictive prosite, pfam or SCOP motif.

The new protein can find application in studying the expression profile of testis-specific genes.

weak similarity to kinesin associated protein KAP3

complete cDNA, complete cds, few EST hits

Sequenced by BMF2

Locus: unknown

Insert length: 2497 bp

Poly A stretch at pos. 2424, polyadenylation signal at pos. 2400

```
1  GGAATCCAAA GAAACAGTTA TGATGGGGGA CTCTATGGTG AAAATAAATG
51 GGATTTATTT AACAAAATCA AATGCTATTT GCCACTTAAA GAGTCACCCA
101 CTTCAGCTAA CTGATGATGG AGGCTTCAGT GAAATAAAGG AGCAAGAAAT
151 GTTCAAAGGA ACAACATCTT TACCATCTCA TCTCAAGAAT GGAGGGGACC
201 AGGGGAAGAG ACATGCGAGG GCCTCATCAT GCCCCAGTAG CTCAGACCTG
251 AGCAGGCTGC AAACCAAAGC AGTCCCAAAA GCTGACCTGC AAGAAGAGGA
301 CGCAGAAATA GAAGTAGACG AAGTCTTTTG GAATACAAGG ATTGTACCGA
351 TTTTGCCTGA ATTAGAAAAG GAAGAAAACA TTGAAACGGT TTGTGCTGCT
401 TGCACACAAC TTATCATATG TTTAGAGGAA GGAAACATGC TTGGAATATA
451 ATTTAAGGGA AGAAGTATTC TCCTGAAGAC CCTGTGTAAA CTAGTTGATG
501 TTGGTTCAGA CTCGCTCAGC CTTAAACTTG CAAAAATAAT TCTAGCACTT
551 AAAGTGAGTA GAAAGAATCT TCTTAATGTC TGCAAACTTA TATTTAAAT
601 TAGCAGGAAT GAGAAGAATG ATTCTTTGAT TCAAAATGAC AGCATTCTGG
651 AATCATTATT GGAGGTACTA AGAAGTGAAAG ACCTGCAAAAC TAACATGGAA
701 GCTTTTTTAT ACTGTATGGG GTCTATAAAG TTCATTCTG GAAATCTGGG
751 ATTTCTTAAT GAAATGATCA GCAAAGGTGC TGTGGAATAA CTGATAAAT
801 TGATAAAACA AATAAATGAG AACATCAAGA AATGGGTAC ATTTTGGCT
851 AATTGCGGCC ACTTGCTAGT CCAGGTGACT GCTACATTGA GAAACTTGGT
901 TGATTCACTA TTAGTAAGAA GTAAGTTCCT AAACATCAGT GCCCTTCCCC
951 AGCTCTGCAC GGCAATGGAA CAGTACAAGG GTGACAAGGA CGTCTGTACC
1001 AATATTGCCA GAATATTCAG CAAACTTACT TCTTACCGTG ACTGCTGCAC
1051 AGCCTTGGCC AGCTATTCCA GATGTTATGC CTTATTCTG AATCTAATTA
1101 ACAAAATCCA GAAGAAGCAG GATTTAGTCG TCCGTGTTGT TTTTATCTTT
1151 GGCAACCTGA CGGCAAAAAA TAACCAGGCT CGTGAACAAAT TTTCAAAAGA
1201 GAAAGGGAGC ATCCAAACTC TGCTGTCATT ATTCCAGACG TTCCATCAGC
1251 TGGATCTGCA TTCCAGAAG CCGGTGGGCC AACGAGGCGA GCAGCACAGG
1301 GCGCAGAGGC CGCCGTCAGA GGCAGAGGAC GTGCTCATCA AGCTGACTCG
1351 TGTGCTGGCC AACATTGCCA TCCACCGGGG CGTGGGCCCG GTGCTGGCCG
1401 CCAACCCGGG GATAGTGGG CTGCTCCTGA CCACGCTGGA ATACAAGTCA
1451 CTTGATGATT GTGAGGAGCT GGTGATCAAT GCTACAGCGA CAATCAACAA
1501 TTTATCTTAC TACCAAGTGA AGAATTCAT AATTCAGAC AAAAAGCTAT
1551 ATATTGCTGA ATTGCTCTTA AAGCTTCTTG TCAGTAACAA CATGGATGGA
1601 ATCCTGGAGG CTGTGCGTGT TTTCGGAAAT CTCTCCAGG ACCATGATGT
1651 CTGCGATTTC ATTGTGCAGA ACAATGTCCA CAGGTTTCATG ATGGCGCTGC
1701 TGGATGCTCA GCATCAGGAT ATCTGCTTTT CTGCTGTGG TGTCTCTCTC
1751 AATCTCACTG TGGATAAAGA CAAGCGTGTG ATCTTGAAG AAGGAGGTGG
1801 CATTAAAAAG TTAGTGGACT GTTTAAGAGA TTTGGTCTCT ACTGATTGGC
1851 AGCTGGCCTG CTTGGTTTGT AAAACTTTAT GGAACCTCAG TGAAAAACATC
1901 ACTAATGCTT CGTCATGTTT TGGAAATGAA GACACCAACA CACTCTTACT
1951 CTTGCTCTCA TCATTTTGTG ATGAAGAACT AGCACTGGAT GGCAGTTTGT
2001 ATCCAGACCT AAAAACTAT CACAAACTCC ATTGGGAAAC AGAATTCAAA
2051 CCGTGGGCAC AGCAGCTTCT AAACCGAATT CAGAGACATC ACACCTTCTCT
2101 GGAACCCCTG CCCATCTCCT CTTTCTAACA TGATGCAGAT TAACAGTAGA
2151 AACGAGAACT CACGCTCTCC TCATTCTTAA GAACGTGTAA CAACGTGAA
2201 CATTTTTTTC AGCATTAACA AATGTGGAAA GTTTTCTAAG AACTGGTTTT
2251 AGTGAGTAGC TGAAGTATTT TTTAAATTA AGCATTCTCT CTTGTTAGGT
2301 ATTATGAAA AATGAATATA CACATTATAT TTCCTGTTGA GAGAAATGTA
2351 AGATGAAAT ATGTGCATT TCAAGTAAAT GACTTTTCT TCTATTCTCT
2401 ATTAACAAT TTAGTCTAG TCTTAAAAA AAAAAAATA AAAAAAATA
2451 AAAAAAATA AAAAAAATA AAAAAAATA AAAAAAATA AAAAAAATA
```

BLAST Results

-----

No BLAST result

Medline entries

No Medline entry

Peptide information for frame 2

ORF from 20 bp to 2125 bp; peptide length: 702  
Category: putative protein

```
1 MMGDSMVKIN GIYLTNSNAI CHLKSHPLQL TDDGGFSEIK EQEMFKGTTS
51 LPSHLKNGGD QGKRHRARASS CPSSSDLSRL QTKAVPKADL QEEDAIEIVD
101 EVFWNTRIVP ILRELEKEEN IETVCAACTQ LHHALEEGNM LGNKFKGRSI
151 LLKTLCKLVD VGSDSLSLKL AKIILALKVS RKNLLNVCKL IPKISRNEKN
201 DSIQNDISIL ESLLLEVRSE DLQTNMEAFI YCMGSIKFIS GNLGFLNEMI
251 SKGAVEILIN LIQINENIK KCGTFLPNSG HLLVQVTATL RNLVDSSLVR
301 SKFLNISALP QLCTAMEQYK GDKDVCTNIA RIFSKLTSYR DCCTALASYS
351 RYALFLNLI NKYQKQDLV VRVVFILGNL TAKNNQAREQ FSKEKGSIQI
401 LLSLFQTFHQ DLHLSQKPVG QRGEQHRAGR PPSEADVLI KLTRVLANIA
451 IHPGVGPVLA ANPGIVGLLL TTLEYKSLDD CEELVINATA TINNLSYYQV
501 KNSIIQDKKL YIAELLKLL VSNMMDGILE AVRFGNLSQ DHDVCDFIQV
551 NNVRFRMMAL LDAQHQDIFC SACGVLLNLT VDKDKRVLK EGGGKIKLVD
601 CLRDLGPTDW QLACLVCCKL WNFSENITNA SSCFGNEDTN TLLLLSSFL
651 DEELALDGSF DPDLKNYHKL HWETEFKPVA QLLNRIQRH HTFLEPLPIP
701 SF
```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKF2phtes3\_14p7, frame 2

TREMBL:MMD367\_1 product: "KAP3B"; Mus musculus mRNA for KAP3B,  
complete cds., N = 2, Score = 97, P = 0.00039

>TREMBL:MMD367\_1 product: "KAP3B"; Mus musculus mRNA for KAP3B, complete  
cds.

Length = 772

HSPs:

Score = 97 (14.6 bits), Expect = 3.9e-04, Sum P(2) = 3.9e-04  
Identities = 45/163 (27%), Positives = 77/163 (47%)

Query: 442 LTRVLANIAIHPGVGPVLAANPGIVGLLLTTLEYKSLDDCEELVINATATINNLSYYQVK 501  
L +++ NI+ H G P VG L + S D+ EE VI T+ NL+ +

Sbjct: 483 LMKMIRNISQHDG--PTKNLFIDYVGDLAAQI---SSDEEEFVIECLGTLANLTIPDL 537

Query: 502 -NSIIQDKKLYIAELLKLLVSNMMDG-ILEAVRVFGNLSQDHDVCDFIQNNVHRFMA 559  
++++ KL + L KL D +LE V + G +S D + ++ + ++

Sbjct: 538 WELVLKEYKL-VPFLKDKLPGAAEDDLVLEVIMIGTVSMDDSCAALLAKSGIIPALIE 596

Query: 560 LLDAQHQDIFCSACGVLL--NLTVDKDKR-VILKEGGGIKLVDCCLRD 604  
LL+AQ +D F C ++ + + R VI+KE L+D + D

Sbjct: 597 LLNAQQEDDEF-VCQIIYVFYQMVFHQATRDVIKETQAPAYLIDLMD 644

Score = 77 (11.6 bits), Expect = 3.9e-04, Sum P(2) = 3.9e-04  
Identities = 42/178 (23%), Positives = 82/178 (46%)

Query: 169 KLAKIILALKVSRKNLLNVCK-LIFKISRNEKNDSLIQNDISILESLLLEVRSEDLQTNME 227  
K K L V ++ LL V L+ ++ + + + ++N +I+ L++ L + N E

Sbjct: 263 KTFKRYQGLVVKQEQLLRVALYLLNLAEDTRTELKMRNKNIVHMLVKALDLD---NFE 318

Query: 228 AFLYCMGSIKFISGNLGFLENEMISKGAVEILINLIKQINENIKKCGTFLPNSGHLLVQVT 287  
+ + +K +S + N+M+ VE L+ +I +E++ L + +

Sbjct: 319 LLILVVSFLKKSIFMENKNDMVEMDIVEKLVKMIPEHEDL-----LNITLR 366

Query: 288 ATLRLNVDSLSVRSKFLNISALPQLCTAM--EQYKGDVCT--NIARI--FSKLTSYRD 341  
L D+ L R+K + + LP+L + E YK +C +I+ F + +Y D

Sbjct: 367 LLLNLSFDITGL-RNKMVQVGLLPRLTALLGNENYK-QIAMCVLYHISMDDRFKSMFAYTD 424

Query: 342 CCTAL 346  
C L  
Sbjct: 425 CIPQL 429

Score = 69 (10.4 bits), Expect = 2.6e+00, Sum P(2) = 9.2e-01  
Identities = 35/146 (23%), Positives = 70/146 (47%)

Query: 512 IAE L L L K L L V S N M D G I L E A V R V F G N L S Q D H D V C D F I V Q N N V H R F M M L L D A Q H Q D I C F S 571  
I +L+K L +N+ ++ V LS + + +V+ ++ ++ ++ +H+D+  
Sbjct: 304 IVHMLVKALDRDNFELLILVVSFLKKLSIFMENKNDVMEDIVEKLVKMIPEHEDLINI 363

Query: 572 ACGVLLNLTVDKDKRVILKEGGGIKKLVDCRLDLPDQW-QLACLVCCKTLWNFSENITNA 630  
+LLNL+ D R + + G + KL L G ++ Q+A +C L++ S +  
Sbjct: 364 TLRLLNLSFDTGLRNKMVQVGLLPKLTALL---GNENYKQIA--MC-VLYHISMD-DRF 416

Query: 631 SSCFGNEDT-NTLLSSFLDEELALD 657  
S F D L+ +L DE + L+  
Sbjct: 417 KSMFAYTDCIPQLMKMLFECSDERIDLE 444

Score = 68 (10.2 bits), Expect = 3.2e-03, Sum P(2) = 3.2e-03  
Identities = 18/58 (31%), Positives = 30/58 (51%)

Query: 190 LIFKISRNEKN-DSLIQND SILES LLEVLRSE-----DLQTNMEAFLYCMGSIKIFISG 241  
LI +++RN N + L+ N++ L +L VLR + +L TN+ +C S G  
Sbjct: 155 LILQLARNPDNLEELLNETALGALARVLRDQKQSVELATNIIYIFCFSSFSHFHG 212

Score = 65 (9.8 bits), Expect = 6.4e+00, Sum P(2) = 1.0e+00  
Identities = 26/122 (21%), Positives = 53/122 (43%)

Query: 283 LVQVTATLRNL---VDSSLVRSKFLNISALPQLCTAMEQYKGDQVCTNIARIFSKLTS 338  
+++ TL NL +D LV ++ +P L ++ +D+ + I S  
Sbjct: 521 VIECLGTLANLTIPLDWELVKEY---KLVPFLKDKLPGAAEDDLVLEV-IMIGTVS 576

Query: 339 YRDCCTALASYSRCYALFLNLINKYQKKQDLVVRVVFILGNLTAKNNQAREQFSKEKCSI 398  
D C AL + S + L+N Q+ + V +++++ + R+ KE +  
Sbjct: 577 MDDSCAALLAKSGIIPALIELLNAQQEDEFVQCIIYVFYQMVF-HQATRDVVIKETQAP 635

Query: 399 QTLSSL 404  
L+ L  
Sbjct: 636 AYLIDL 641

Score = 65 (9.8 bits), Expect = 6.4e+00, Sum P(2) = 1.0e+00  
Identities = 44/177 (24%), Positives = 79/177 (44%)

Query: 481 CE-ELVINATATIN-NLSYYQ-VKNSIIQDKLYIAELLKLLVSNMMDGILEAVRVFGN 537  
CE E ++N T + NLS+ ++N ++Q + + L LL + N I A+ V  
Sbjct: 355 CEHEDLLNITRLRLNLSFDTGLRNKMVQ---VGLLPKLTALLGNENYKQI--AMCVLYH 409

Query: 538 LSQDHDVCD-FIVQNNVHFRFMALLDAQHQDICFSACGVLLNLTVDKDKRVILKEGGGIK 596  
+S D F + + + H L + + I L +K ++ EG +K  
Sbjct: 410 ISHDDRFSKMFAYTDCIPQLMKMLFECSDERIDLELISFCINLAANKRNVQLICEGNGLK 469

Query: 597 KLVDCLRDLPDQWQLACLVCCKTLWNFSENITNASSCFGNEDTNTLLSSFLDEELAL 656  
L+ R L D L+ K + N S++ + F + L +SS +EE +  
Sbjct: 470 MLMK--RALKLKD---PLMKMIRNISQHDGPTKNLF-IDYVGDLAAQISSDEEEFVI 522

Query: 657 D 657  
+  
Sbjct: 523 E 523

Score = 61 (9.2 bits), Expect = 1.6e-02, Sum P(2) = 1.6e-02  
Identities = 20/66 (30%), Positives = 34/66 (51%)

Query: 304 LNISALPQLCTAM-EQYKGDQVCTNIARIFSKLTSYRDCCTALASYSRCYALFLNLINK 362  
LN +AL L + E +K ++ TNI IF +S+ + + Y + AL +N+I+  
Sbjct: 171 LNETALGALARVLRDQKQSVELATNIIYIFCFSSFSHFHGLITHY-KIGALCMNIIDH 229

Query: 363 YQKKQDL 369  
K+ +L  
Sbjct: 230 ELARHEL 236

Pedant information for DKFZphtes3\_14p7, frame 2

Report for DKFZphtes3\_14p7.2

[LENGTH] 708  
[MW] 79266.35  
[pI] 6.57

```

(FUNCAT)      30.25 vacuolar and lysosomal organization [S. cerevisiae, YEL013w] 3e-04
(FUNCAT)      06.04 protein targeting, sorting and translocation [S. cerevisiae, YEL013w]
3e-04
(FUNCAT)      09.25 vacuolar and lysosomal biogenesis [S. cerevisiae, YEL013w] 3e-04
(BLOCKS)      BL00923F Aspartate and glutamate racemases proteins
(BLOCKS)      BL00288B Tissue inhibitors of metalloproteinases proteins
(PSITE)        MYRISTYL 9
(PSITE)        AMIDATION 1
(PSITE)        CK2_PHOSPHO_SITE 12
(PSITE)        PKC_PHOSPHO_SITE 7
(PSITE)        ASN_GLYCOSYLATION 11
(KW)           Alpha_Beta
(KW)           LOW_COMPLEXITY 7.49 %

```

```

SEQ      ESKETVMGDSMVKINGIYLTKSNAICHLKSHPLQLTDDGGFSEIKEQEMFKGTTSLPSH
SEG      .....
PRD      cccceeecccccccccccccccccccccccccccccccccccccccccccccccccccccccc

```

```

SEQ      LKNGGDQGRKHARASSCPSSDLSRLQTKAVPKADLQEDAEIEVDEVFWNTRIVPILRE
SEG      .....
PRD      cccceeecccccccccccccccccccccccccccccccccccccccccccccccccccccccc

```

```

SEQ      LEKEENIETVCAACTQLHHALEEGNMLGNFKGRSILLKTLCKLVDSLSLKLAKII
SEG      .....
PRD      hhhhhcccccccccccccccccccccccccccccccccccccccccccccccccccccccc

```

```

SEQ      LALKVSRKNLLNVCKLIFKISRNEKNDLIQNDLSILESLEVLSEDLQTNMEAFLYCMG
SEG      .....
PRD      hhhhhcccccccccccccccccccccccccccccccccccccccccccccccccccccccc

```

```

SEQ      SIKFISGNLGFLENMISKGAVEILINLIKQINENIKKCGTFLPNSGHLVQVATLRNLV
SEG      .....
PRD      cccceeecccccccccccccccccccccccccccccccccccccccccccccccccccccccc

```

```

SEQ      DSSLVRSKFLNISALPQLCTAMEQYKGDQDVCTNIARIFSKLTSYRDCCTALASYSRCYA
SEG      .....
PRD      cccceeecccccccccccccccccccccccccccccccccccccccccccccccccccccccc

```

```

SEQ      LFLNLINKYQKQDLVVRVVFILGNLTAKNNQAREQFSKERGSIQTLTSLFQTFHQLDLH
SEG      .....
PRD      hhhhhcccccccccccccccccccccccccccccccccccccccccccccccccccccccc

```

```

SEQ      SQKPVGQRGEQHRAPPPSEADVLIKLTRLVLANIAIHGPGVPLAANPGIVGLLLTLE
SEG      .....
PRD      cccceeecccccccccccccccccccccccccccccccccccccccccccccccccccccccc

```

```

SEQ      YKSLDDCEELVINATATINLSYYQVKNIIQDKKLYIAELLKLLVSNMMDGILEAVRV
SEG      .....
PRD      hhhcccccccccccccccccccccccccccccccccccccccccccccccccccccccc

```

```

SEQ      FGNSLQDHDVDFIVQNNVHRFMALLDAQHQCICFSACGVLLNLTVDKDKRVILKEGGG
SEG      .....
PRD      cccceeecccccccccccccccccccccccccccccccccccccccccccccccccccccccc

```

```

SEQ      IKKLVDCLDLGPTDQLACLVCCKLWNFSEINASSCFGNEDTNTLLLSFLDEEL
SEG      .....
PRD      hhhhhcccccccccccccccccccccccccccccccccccccccccccccccccccccccc

```

```

SEQ      ALDGSFDPDLKNYHKLHWETEFKPAQQLLNRIQRHHTFLEPLPIPSF
SEG      .....
PRD      hhhcccccccccccccccccccccccccccccccccccccccccccccccccccccccc

```

#### Prosites for DKFZphtes3\_14p7.2

PS00001	206->210	ASN_GLYCOSYLATION	PDOC00001
PS00001	212->216	ASN_GLYCOSYLATION	PDOC00001
PS00001	311->315	ASN_GLYCOSYLATION	PDOC00001
PS00001	385->389	ASN_GLYCOSYLATION	PDOC00001
PS00001	493->497	ASN_GLYCOSYLATION	PDOC00001
PS00001	500->504	ASN_GLYCOSYLATION	PDOC00001
PS00001	543->547	ASN_GLYCOSYLATION	PDOC00001
PS00001	584->588	ASN_GLYCOSYLATION	PDOC00001
PS00001	628->632	ASN_GLYCOSYLATION	PDOC00001
PS00001	632->636	ASN_GLYCOSYLATION	PDOC00001
PS00001	635->639	ASN_GLYCOSYLATION	PDOC00001
PS00005	173->176	PKC_PHOSPHO_SITE	PDOC00005
PS00005	186->189	PKC_PHOSPHO_SITE	PDOC00005
PS00005	241->244	PKC_PHOSPHO_SITE	PDOC00005



WO 01/12659

PCT/IB00/01496

PS00005	295->298	PKC_PHOSPHO_SITE	PDOC00005
PS00005	344->347	PKC_PHOSPHO_SITE	PDOC00005
PS00005	387->390	PKC_PHOSPHO_SITE	PDOC00005
PS00005	421->424	PKC_PHOSPHO_SITE	PDOC00005
PS00006	79->83	CK2_PHOSPHO_SITE	PDOC00006
PS00006	201->205	CK2_PHOSPHO_SITE	PDOC00006
PS00006	214->218	CK2_PHOSPHO_SITE	PDOC00006
PS00006	218->222	CK2_PHOSPHO_SITE	PDOC00006
PS00006	230->234	CK2_PHOSPHO_SITE	PDOC00006
PS00006	320->324	CK2_PHOSPHO_SITE	PDOC00006
PS00006	344->348	CK2_PHOSPHO_SITE	PDOC00006
PS00006	439->443	CK2_PHOSPHO_SITE	PDOC00006
PS00006	477->481	CK2_PHOSPHO_SITE	PDOC00006
PS00006	483->487	CK2_PHOSPHO_SITE	PDOC00006
PS00006	654->658	CK2_PHOSPHO_SITE	PDOC00006
PS00006	698->702	CK2_PHOSPHO_SITE	PDOC00006
PS00008	17->23	MYRISTYL	PDOC00008
PS00008	64->70	MYRISTYL	PDOC00008
PS00008	144->150	MYRISTYL	PDOC00008
PS00008	384->390	MYRISTYL	PDOC00008
PS00008	402->408	MYRISTYL	PDOC00008
PS00008	473->479	MYRISTYL	PDOC00008
PS00008	533->539	MYRISTYL	PDOC00008
PS00008	580->586	MYRISTYL	PDOC00008
PS00008	641->647	MYRISTYL	PDOC00008
PS00009	67->71	AMIDATION	PDOC00009

(No Pfam data available for DKF2phtes3\_14p7.2)

DKF2phtes3\_15a13

-----

group: testes derived

DKF2phtes3\_15a13 encodes a novel 387 amino acid protein with weak similarity to *S.cerevisiae* Hop1.

No informative BLAST results; No predictive prosite, pfam or SCOP motif.

The new protein can find application in studying the expression profile of testis-specific genes.

similarity to *S.cerevisiae* Hop1

complete cDNA, complete cds, potential start codon at Bp 116, 3 EST

hits

*S.cerevisiae* Hop1p is a meiosis-specific protein

Sequenced by GBF

Locus: unknown

Insert length: 1848 bp

Poly A stretch at pos. 1766, no polyadenylation signal found

```
1  GGAAGCGCA  TCGCGTCGG  GCACAGCGG  TGCAGCCTCG  TGCAGCTCTT
51  CTGGTCTCCG  GCGCCCGCCC  CTCAGACGTA  ATGTTGAATT  AAAGAAAAATA
101  CTTTATCAGA  AGAAGATGGC  CACTGCCAG  TTGCAGAGGA  CTCCCATGAG
151  TGCACCTGGTA  TTTCCCAATA  AGATATCAAC  TGAACACCAG  TCTTTGGTGT
201  TAGTGAAGAG  GCTTCTAGCA  GTTTCAGTAT  CCTGTATCAC  GTATTTGAGG
251  GGAATATTCC  CAGAATGCGC  TTATGGAACA  AGATATCTAG  ATGATCTTTG
301  TGCAAAATA  CTGAGAGAA  ATAAAAATTG  CCCAGGATCT  ACACAGTTAG
351  TGAATGGAT  GCTAGGATGT  TATGATGCTT  TACAGAAAAA  ATATGTATAC
401  ACAACCCAG  AAGATCCTCA  GACAATTCA  GAATGTTACC  AATTCAAATT
451  CAAATACACC  AATAATGGAC  CACTCATGGA  CTTATAAGT  AAAAACCAAA
501  GCAACGAATC  TAGCATGTTG  TCTACTGACA  CCAAGAAAGC  AAGCATTCTC
551  CTCATTGCGA  AGATTTATAT  CCTAATGCAA  AATCTGGGGC  CTTTACCTAA
601  TGATGTTTGT  TTGACCATGA  AACTTTTTTA  CTATGATGAA  GTTACACCCC
651  CAGATTACCA  GCCTCCCGGT  TTTAAGGATG  GTGATTGTGA  AGGAGTTATA
701  TTTGAAGGGG  AACCTATGTA  TTTAAATGTG  GGAGAGTCT  CAACACCTTT
751  TCACATCTTC  AAAGTAAAG  TGACCACTGA  GAGAGAACGA  ATGGAAAAATA
801  TTGACTCAAC  TATACTATCA  CCAAAACAAA  TAAAAACACC  ATTTCAAAAA
851  ATCCTGAGGG  ACAAGATGT  AGAAGATGAA  CAGGAGCATT  ATACAAGTGA
901  TGATTTGGAC  ATTGAAACTA  AAATGGAAGA  ACAGGAAAAA  AACCTGCAT
951  CTTCTGAAC  TGAAGAACCA  AGTTTAGTTT  GTGAGGAAGA  TGAATTATG
1001  AGGCTCTAA  AAAGTCCAGA  TCTTCTATT  TCTCATTCTC  AGGTTGAGCA
1051  GTTAGTCAAT  AAAACATCTG  AACTTGATAT  GTCTGAAAGC  AAAACAAGAA
1101  GTGGAAAAGT  CTTTCAGAA  AAAATGGCAA  ATGGAAATCA  ACCAGTAAAA
1151  TCTTCCAAAG  AAAATCGGAA  GAGAAGTCAA  CATGAATCTG  GGAGAAATAGT
1201  CCTCCATCAC  TTTGATTCTT  CTAGTCAAGA  GTCAGTGCCA  AAAAGGAGAA
1251  AGTTTAGTGA  ACCAAAGGAA  CATATATAAA  AATTATTTTT  GTTCTGCAGG
1301  CTTCGAGAGT  TCTTCTCACC  ATTTAACTG  AAGGACCCTA  TATTATATTT
1351  CCCTAACTCT  GAAGATGTAT  ATGTAGTTTA  AAGCAGTTTG  TACATAAAA
1401  CTAAGTTTTT  GGCTGACTGT  CATATTGTGG  TCCTTAATCT  TGAGATAAAT
1451  CCAATAGAAC  TTTTGAATA  AAGCAAAAGT  ACAAATGTCA  TAATTGATTC
1501  GGTAATAAGT  AAAATTTCAA  AATTGATTTT  GTTCATTACC  TACTTAATAT
1551  TTCCTTTAAA  TATATACTAA  CTGTTAAGGC  CCTCTAATGC  CATTTTTCTA
1601  AACAGTAATG  TTTACTTTGG  TATTAATAAT  TGGTATGGAT  TCACTTTTTA
1651  CTTATGTTAA  AATTATACCA  TTTAACTGGC  TCTTTGTGCA  TTGTGCTGTT
1701  ATTAACACAA  TGTTCTTCAA  TATTTTGACA  TAATGTATTA  ACATTTTAAT
1751  ATATAATGTA  CAATTTAAAA  AAAAAAAAAA  AAAAAAAAAA  AAAAAAAGG
1801  GCGCGCGCT  CTAGAGGATC  CAAGCTTACG  TACAAAAAAA  AAAAAAAGG
```

## BLAST Results

-----

No BLAST result

## Medline entries

-----

No Medline entry

## Peptide information for frame 2

ORF from 116 bp to 1276 bp; peptide length: 387  
Category: similarity to known protein

```
1 MATAQLQRTF MSALVFPNKI STEHQSLVLV KRLLAHSVSC ITYLRGIFPE
51 CAYGTRYLDD LCVKILREDK NCPGSQLVK WMLGCDALQ KKYVYTNPED
101 PQTISECYQF KFKYTNNGL MDFISKNQSN ESSMLSTDK KASILLIRKI
151 YILMQNLGPL PNDVCLTMKL FYYDEVTPPD YQPPGFKDGD CEGVIFEGEP
201 MYLNVGEVST PFHIFKVKVT TERERENID STILSPKQIK TPFQKILRDK
251 DVEDEQEHYT SDDLDIETKM EEQKNPASS ELEEPSLVCE EDEIMRSKES
301 PDLSSHSQV EQLVNKTSSEL DMSKSTRSG KVFQNMANG NQPVKSSKEN
351 RKRSQHESGR IVLHHFDSST QESVPKRRKF SEPKEHI
```

## BLASTP hits

No BLASTP hits available

## Alert BLASTP hits for DKFZphtes3\_15a13, frame 2

TREMBL:ATAC2130\_3 product: "F1N21.3"; The sequence of BAC F1N21 from Arabidopsis thaliana chromosome 1, complete sequence., N = 1, Score = 274, P = 5.7e-22

TREMBL:SC9877\_9 gene: "hop1"; S.cerevisiae chromosome IX cosmid 9877., N = 2, Score = 126, P = 7.1e-09

PIR:A34691 meiosis-specific protein HOP1 - yeast (Saccharomyces cerevisiae), N = 2, Score = 126, P = 7.8e-08

>TREMBL:ATAC2130\_3 product: "F1N21.3"; The sequence of BAC F1N21 from Arabidopsis thaliana chromosome 1, complete sequence.  
Length = 562

## HSPs:

Score = 274 (41.1 bits), Expect = 5.7e-22, P = 5.7e-22  
Identities = 84/290 (28%), Positives = 145/290 (50%)

```
Query: 22 TEHQSLVLVKRLLAHSVSCITYLRGIFPECAYGTRYLDDLCVKILREDKNCPGSQLVKW 81
TE SL+L + LL +++ I+Y+RG+FPE + + + L +KI + S +L+ W
Sbjct: 11 TEQDSLLLRNLLRIAIFNISYIRGLFPEKYFNKSVPALDMKIKLMPMDAESRRLIDW 70

Query: 82 M-LGCDALQKKYVYT-----NPEDPQTISECYQFKFYTNNGP--LMDFSK--NQSN 130
M G YDALQ+KY+ T D I E Y F F Y+++ +M I++ N+ N
Sbjct: 71 MEKGVYDALQRKYLKTLMFISICETVDGPMIEE-YSFSFSYSDSDSQVMMNINRTGNKKN 129

Query: 131 ESSMLST-----DTKKASILLIRKIYILMQNLGFLPNDVCLTMKLFYYDEVTPPDYQPP 184
ST + + + +R + LM+ L +P++ + MKL YYD+VTPPDY+PP
Sbjct: 130 GGIFNSTADITPNQMRSSACKMVRTLVLQMLTLDKMPDERTIVMKLLYYDVTTPDYEP 189

Query: 185 GFKD--GDCEGVIFEGEPMYLNVGEVSTPFHIFKVKVT-----ERERENIDSTILS 235
F+ D ++ P+ + +G V++ + +KV + E + M++ D +
Sbjct: 190 FFRGCTEAEQYVWTKNPLRMEIGNVNSKHLVLTLLKVSVDLPCEDEDDMQD-DGKSIG 248

Query: 236 PQQIKTPFQKILRDKDVEDEQEHY-----TSDDLDIETKMEEQKNPASSE 281
P + Q D ++ QE+ DD D E ++ +PA +E
Sbjct: 249 PDSVHDD-QPSDSSEISQTQENQFIVAPVEKQDDDDGEVDDNTQDPAENE 300
```

## Pedant information for DKFZphtes3\_15a13, frame 2

## Report for DKFZphtes3\_15a13.2

```
[LENGTH] 387
[MW] 44417.64
[pI] 5.57
[HOMOL] TREMBL:ATAC2130_3 product: "F1N21.3"; The sequence of BAC F1N21 from
Arabidopsis thaliana chromosome 1, complete sequence. 9e-23
[FUNCAT] 09.13 biogenesis of chromosome structure [S. cerevisiae, YIL072w] 7e-11
[FUNCAT] 03.19 recombination and dna repair [S. cerevisiae, YIL072w] 7e-11
[FUNCAT] 03.13 meiosis [S. cerevisiae, YIL072w] 7e-11
[FUNCAT] 30.10 nuclear organization [S. cerevisiae, YIL072w] 7e-11
[PIRKW] nucleus 2e-09
[PIRKW] zinc finger 2e-09
```

```
[PIRKW]      DNA binding 2e-09
[PROSITE]    MYRISTYL      1
[PROSITE]    CAMP_PHOSPHO_SITE      3
[PROSITE]    CK2_PHOSPHO_SITE      12
[PROSITE]    PKC_PHOSPHO_SITE      7
[PROSITE]    ASN_GLYCOSYLATION      3
[KW]         Alpha_Beta

SEQ  MATAQLQRTPMASLVFPNKISTEHQSLVLVKRLAVSVSCITYLRGIFPECAYGTRYLDD
PRD  ccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc

SEQ  LCVKILREDKNCPCGSQLVWMLGQYDALQKKYVYTNPEDPQTISECYQFKKYTNNGPL
PRD  hhhhhhhcccccccccccccccccccccccccccccccccccccccccccccccccccc

SEQ  MDFISKQSNESMLSTDTKKASILLIRKIYILMQNLGPLPNDVCLTMKLFYYDEVTPPD
PRD  eeeeecccccccccccccccccccccccccccccccccccccccccccccccccccccccc

SEQ  YQPPGFKDGDCEGVIFEQEPMYLNVGEVSTPFHIFKVKVTTTEREREMIDSTILSPKQIK
PRD  ccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc

SEQ  TPFQKILRDKDVEDEQEYHTSDDLOICTKMEEQEKNPASSELEEPSLVCEEDEIMRSKES
PRD  hhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh

SEQ  PDLISHSQVEQLVNKTSLEDMSESKTRSGKVFQNKMANGNQPVKSSKENRKRSQHESGR
PRD  cccccchhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh

SEQ  IVLHHFDSSSQESVPKRKRFSEPKHEI
PRD  eeeeecccccccccccccccccccccccc
```

Prosite for DKF2phtes3\_15a13.2

PS00001	127->131	ASN_GLYCOSYLATION	PDOC00001
PS00001	130->134	ASN_GLYCOSYLATION	PDOC00001
PS00001	315->319	ASN_GLYCOSYLATION	PDOC00001
PS00004	140->144	CAMP_PHOSPHO_SITE	PDOC00004
PS00004	351->355	CAMP_PHOSPHO_SITE	PDOC00004
PS00004	378->382	CAMP_PHOSPHO_SITE	PDOC00004
PS00005	139->142	PKC_PHOSPHO_SITE	PDOC00005
PS00005	167->170	PKC_PHOSPHO_SITE	PDOC00005
PS00005	221->224	PKC_PHOSPHO_SITE	PDOC00005
PS00005	235->238	PKC_PHOSPHO_SITE	PDOC00005
PS00005	329->332	PKC_PHOSPHO_SITE	PDOC00005
PS00005	346->349	PKC_PHOSPHO_SITE	PDOC00005
PS00005	358->361	PKC_PHOSPHO_SITE	PDOC00005
PS00006	96->100	CK2_PHOSPHO_SITE	PDOC00006
PS00006	103->107	CK2_PHOSPHO_SITE	PDOC00006
PS00006	177->181	CK2_PHOSPHO_SITE	PDOC00006
PS00006	221->225	CK2_PHOSPHO_SITE	PDOC00006
PS00006	260->264	CK2_PHOSPHO_SITE	PDOC00006
PS00006	268->272	CK2_PHOSPHO_SITE	PDOC00006
PS00006	280->284	CK2_PHOSPHO_SITE	PDOC00006
PS00006	308->312	CK2_PHOSPHO_SITE	PDOC00006
PS00006	318->322	CK2_PHOSPHO_SITE	PDOC00006
PS00006	346->350	CK2_PHOSPHO_SITE	PDOC00006
PS00006	354->358	CK2_PHOSPHO_SITE	PDOC00006
PS00006	369->373	CK2_PHOSPHO_SITE	PDOC00006
PS00008	84->90	MYRISTYL	PDOC00008

(No Pfam data available for DKF2phtes3\_15a13.2)

DKF2phtes3\_15c24

-----

group: metabolism

DKF2phtes3\_15c24 encodes a novel 404 amino acid protein with strong similarity to 2-hydroxyacid dehydrogenases.

The novel protein contains a D-isomer specific 2-hydroxyacid dehydrogenases signature. Proteins with such a signature have similar enzymatic activities: D-lactate dehydrogenase (EC 1.1.1.28), catalyzes the reduction of D-lactate to pyruvate. D-glycerate dehydrogenase (EC 1.1.1.29) catalyzes the reduction of hydroxyypyruvate to glycerate. 3-phosphoglycerate dehydrogenase (EC 1.1.1.95), catalyzes the oxidation of D-3-phosphoglycerate to 3-phosphohydroxyypyruvate. Therefore the novel protein is a new 2-hydroxyacid dehydrogenase.

The new protein can find application in modulation of 2-hydroxyacid dehydrogenases-dependent pathways and as a new enzyme for biotechnologic production processes.

strong similarity to C.elegans T03F1.1

potential start at Bp 55 matches kozak consensus PyCCatgG

Sequenced by GBF

Locus: unknown

Insert length: 1956 bp

Poly A stretch at pos. 1929, polyadenylation signal at pos. 1903

```
1 CGAAGGCGGC GGCAGAGGCC CGGGCTGGGA GCGTTGGCGG CCGGAGTCCC
51 AGCCATGGCG GAGTCTGTGG AGCGCCTGCA GCAGCGGGTC CAGGAGCTGG
101 AGCGGGAACT TGCCCAGGAG AGGAGTCTGC AGGTCCCAGG GAGCGGCGAC
151 GGAGGGGGCG GCCGGGTCCG CATCGAGAAG ATGAGCTCAG AGGTGGTGGA
201 TTCGAATCCC TACAGCCGCT TGATGGCATT GAAACGAATG GGAATTGTAA
251 GCGACTATGA GAAAATCCGT ACCTTTGCCG TAGCAATAGT AGGTGTTGGT
301 GGAGTAGGTA GTGTGACTGC TGAATGCTG ACAAGATGTG GCATTGGTAA
351 GTTGCTACTC TTTGATTATG ACAAGGTGGA ACTAGCCAAT ATGAATAGAC
401 TTTCTTCCA ACCTCATCAA GCAGGATTAA GTAAAGTTCA AGCAGCAGAA
451 CATACTCTGA GGAACATTAA TCCTGATGTT CTTTTGAAG TACACAACTA
501 TAATATAACC ACAGTGGAAA ACTTTCACAA TTTCATGGAT AGAATAAGTA
551 ATGGTGGGTT AGAAGAAGGA AAACCTGTTG ATCTAGTTCT TAGCTGTGTG
601 GACAATTTTG AAGCTCGAAT GACAATAAAT ACAGCTTTGA ATGAAGTTGG
651 ACAAAACATG ATGGAATCTG GGGTCAGTGA AAATGCAGTT TCAGGGCATA
701 TACAGCTTAT AATTCCTGGA GAATCTGCTT GTTTTGCCTG TGCTCCACCA
751 CTTGTAGTTG CTGCAATATG TGATGAAAAG ACTCTGAAAC GAGAAGGTGT
801 TTTGTGACCC AGTCTTCCTA CCACATATGG TGTGGTTGCT GGGATCTTAG
851 TACAAAACGT GTTAAAGTTT CTGTTAAATT TTGGTACTGT TAGTTTTTAC
901 CTTGGATACA ATGCAATGCA GGATTTTTTT CCTACTATGT CCATGAAGCC
951 AAATCCTCAG TGTGATGACA GAAATTGCAG GAAGCAGCAG GAGGAATATA
1001 AGAAAAAGGT AGCAGCACTG CCTAAACAG AGGTTATACA AGAAGAGGAA
1051 GAGATAATCC ATGAAGATAA TGAATGGGGT ATTGAGCTGG TATCTGAGGT
1101 TTCAGAAAGG GAACTGAAAA ATTTTTCAGG TCCAGTTCCA GACTTACCTG
1151 AAGGAATTAC AGTGGCATAC ACAATTCCAA AAAAGCAAGA AGATTCTGTC
1201 ACTGAGTTAA CAGTGGGAAGA TTCTGGTGAA AGCTTGGGAA ACCTCATGGC
1251 CAAAATGAAG AATATGTAGA TAATGGACTG GGATATATTG TATTTCTCAT
1301 GTTAAAGCCT CTTCCCTTGA AATTAAAAAA AAATTTTAAC TGATAAAACT
1351 TAGGGCAACA TTAATTAAATG TATATTCTTA CCTGAATTGT TATACTTTTT
1401 GAAAATCCTG TGACTTGCCCT GTTCTCCCCC GCTCCAACGA AATCATTAAAC
1451 TCTCCTAAAA TGTGTTTCAT TCTAGTAAGA AAACCTCAAA GGATATTGTA
1501 GGATATAAAT CTTACTTGAA AACATAGCTG TTGAAATGTT TTGGCCTTTT
1551 GGAGTGGGGG AAGGACAAAT CTGATCCTGT AATCTTTTTC TTTCAGTAA
1601 TCCCTTGTGT CTGTTGCATG AGGACATGGA CAATAAAGTA GTATATGATC
1651 CTCAGATACA GGGAGAAGGA CAAGGCATAC AGCTTATTGA TTAGAGCTGG
1701 CAAGCATCTG CTCATTATGT TTGGAATTGC TTCTATAAG AAAATTGCC
1751 ACTACTACTA ACTTGATCAA CAATGAATTC AAAATAGTTA ACCTATGAAA
1801 TAACATCCTC TCAAAATGTT GCTGATGAAG TACAAGTTGA AATGTAGTTA
1851 TTGGAAAAGT CTGTAACCTG TGGATCATAT ATATTCAAAG TGAGACAAAG
1901 GCAATAAAAA AGCAGCTATT TTCATGAATA GACAAAAAAA AAAAAAAA
1951 AAAAAG
```

## BLAST Results

-----

No BLAST result

## Medline entries

No Medline entry

## Peptide information for frame 1

ORF from 55 bp to 1266 bp; peptide length: 404  
Category: similarity to unknown protein  
Classification: Metabolism  
Prosit motifs: D\_2\_HYDROXYACID\_DH\_1 (76-105)

```
1 MAESVERLQQ RVQELERELA QERSLQVPRS GGGGGGRVRI EKMSSEVVDS
51 NPYSRLMALK RMGIVSDYEK IRTFAVAIVG VGGVGSVTAE MLTRCGIGKL
101 LLFDYDKVEL ANMNRLLFFQ HQAGLSKVQA AEHTLRNINP DVLFEVHNIN
151 ITTVENFQHF MDRISNGGLE EGKPVDLVLS CVDNFEARMT INTACNELGQ
201 TWMSGVSEN AVSGHIQLII PGESACFACA PPLVVAANID ETKLKREGVC
251 AASLPTTMGV VAGILVQNVL KFLNFGTVS FYLGYNAMQD FFPMTSMKPN
301 PQCDNRNCRK QEEYKKKVA ALPKQEVQEE EEEIHEDNE WGIELVSEVS
351 EELKNFSGP VPDLPPEGITV AYTIPKKQED SVTELTVEDS GESLEDLMAR
401 MKNM
```

## BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphtes3\_15c24, frame 1

TREMBL:CEUT03F1\_11 gene: "T03F1.1"; Caenorhabditis elegans cosmid  
T03F1., N = 1, Score = 1204, P = 1.9e-122

TREMBL:ATAC98\_3 gene: "YUP8H12.3"; Arabidopsis thaliana chromosome 1  
YAC yUP8H12 complete sequence., N = 1, Score = 733, P = 1.5e-72

PIR:A69319 thiamine biosynthesis protein (thiF) homolog - Archaeoglobus  
fulgidus, N = 1, Score = 218, P = 1.8e-17

TREMBL:AF022796.4 gene: "moeB"; product: "MoeB"; Staphylococcus  
carneus molybdenum cofactor biosynthetic gene cluster, complete  
sequence., N = 1, Score = 220, P = 3.7e-16

>TREMBL:CEUT03F1\_11 gene: "T03F1.1"; Caenorhabditis elegans cosmid T03F1.  
Length = 419

## HSPs:

Score = 1204 (180.6 bits), Expect = 1.9e-122, P = 1.9e-122  
Identities = 241/367 (65%), Positives = 293/367 (79%)

```
Query: 37 RVRIEKMSSEVVDSNPYSRLMALKRMGIVSDYEKIRTFVAIVGVGGVGSVTAEMLTRCG 96
      R +IEK+S+EVVDSNPYSRLNAL+RMGIV++YE+IR VA+VGVGGVGSV AEMLTRCG
Sbjct: 48 RQKIEKLSAEVVDSNPYSRLNALQRMGIVNEYERIREKTVAVGVGGVGSVVAEMLTRCG 107

Query: 97 IGKLLLFDDYDKVELANMNRLLFFQHQAGLSKVQAAEHTLRNINPDVLFVHNINNTTVEN 156
      IGKL+LFDDYDKVE+ANMNRLLF+QP+QAGLSKV+AA TL ++NPDV EVHN+NITT++N
Sbjct: 108 IGKLLLFDDYDKVEIANMNRLLFYQPNQAGLSKVEAARDTLIHVNPVQIEVHNFNITMDN 167

Query: 157 FQHFMDRISNGGLEEGKPVDLVLSVDNFEARMTINTACNELGQTMMSGVSENNAVSGHI 216
      F F++RI G L +GK +DLVLSVDNFEARM+N ACNE Q WMESGVSENNAVSGHI
Sbjct: 168 FDTFVNRIKGSLLTDGK-IDLVLSVDNFEARMVNMACEENQIWMESGVSENNAVSGHI 226

Query: 217 QLIIPGESACFACAPPLVVAANIDEKTLKREGVCAASLPTTMGVVAGILVQNVLKFLNLF 276
      Q I PG++ACFAC PPLVVA+ IDE+TLKR+GVCAASLPTTM VVAG LV N LK+LLNF
Sbjct: 227 QYIEPGKTACFACVPLVVASGIDERTLKRDGVCAASLPTTMVAVAGFLVMNTLKYLNF 286

Query: 277 GTVSFYLGYNAMQDFFPTMSMKPNPQCDNRNCRKQEEYKKKVAALPKQ-EV-IQEEEEEI 334
      G VS Y+GYNA+ DFFP S+KPNP CDD +C ++Q+EY++KVA P EV +EEE +
Sbjct: 287 GEVSQYVGYNALSDFPRDSIKPNPYCDDSHCLQRQKEYEEKVANQPVDLEVEVPPEETV 346

Query: 335 IHEDNEWGIELVSEVSEELKNFSGPVPDLPPEGITVAYTIIPKKQEDSVTELTVEDSGESL 394
      +HEDNEWGIELV+E SE + S + G+ AY P K+ D+ TEL+ + +
Sbjct: 347 VHEDNEWGIELVNE-SEPSAEQSSSL--NAGTGLKFAYE-PIKR-DAQTELSPAQA--AT 399

Query: 395 EDLMKMKMKN 403
```



DKFZphtes3\_15c6

group: transmembrane protein

DKFZphtes3\_15c6 encodes a novel 118 amino acid protein without similarity to known proteins.

The novel protein contains 1 transmembrane region.

No informative BLAST results; No predictive prosite, pfam or SCOP motif.

The new protein can find application in studying the expression profile of testis-specific genes and as a new marker for testicular cells.

unknown

complete cDNA, complete cds, EST hits

Sequenced by GBF

Locus: unknown

Insert length: 1283 bp

Poly A stretch at pos. 1264, no polyadenylation signal found

```
1 GAGACACTGA GCCCGAGAC AGTGAGTGGT GGCTCACTG CTCTGCCCCG
51 CACCCGTGCA CCTCCACTTT GCCTTGTGG AAGTGACCCA GCCCCTCCC
101 CTTCCTTCT CCCACCTGTT CCCAGGACT CACCCAGCC CTTGCCTGCC
151 CCTGAGGAAG AAGAGGCACT CACCACTGAG GACTTTGAGT TGCTGGATCA
201 GGGGGAGCTG GAGCAGCTGA ATGCAGAGCT GGGCTTGGAG CCAGAGACAC
251 CGCCAAAACC CCTGATGCT CCACCCCTGG GGCCCGACAT CCATTCTCTG
301 GTACAGTCAG ACCAAGAAGC TCAGGCCGTG GCAGAGCCAT GAGCCAGCCG
351 TTGAGGAAGG AGCTGCAGGC ACAGTAGGGC TTCTGGCTA GGAGTGTTC
401 TGTTTCCTCC TTTGCCTACC ACTCTGGGGT GGGCCAGTGT GTGGGGAAGC
451 TGGCTGTCGG ATGGTAGCTA TTCCACCTC TGCTGCCTG CTTGCCCTGT
501 GTCTGGGCA TGGTGAGTA CCTGTGCCA GGATTGGTTT TAAATTTGTA
551 AATAATTTTC CATTTGGGTT AGTGGATGTG AACAGGGCTA GGAAGTCCT
601 TCCCACAGCC TGCCTTGCC TCCCTGCCTC ATCTCTATT TCATTCCACT
651 ATGCCCCAAG CCTGGTGGT CTGGCCCTTT CTTTTCTC CTATCCTCAG
701 GGACCTGTGC TGCTCTGCCC TCATGTCCCA CTGGGTGTT TAGTTAGGCG
751 ACTTTATAAT TTTCTCTTG TCTTGTTTC CTTCTGCTT TATTTCCCTG
801 CTGTGTCCTG TCCTTAGCAG CTCACCCCA TCCTTGCCA GCTCCTCCTA
851 TCCCGTGGGC ACTGGCCAAG CTTTAGGGAG GCTCCTGGT TGGGAAGTAA
901 AGAGTAAACC TGGGGCAGTG GGTGAGGCCA GTAGTTACAC TCTTAGGTCA
951 CTGTAGTCTG TGTAACTTC ACTGCATCCT TGCCCATTC AGCCCGGCT
1001 TTCATGATGC AGGAGAGCAG GGATCCCGCA GTACATGGCG CCAGCACTGG
1051 AGTTGGTGAG CATGTGCTCT CTCTTGAGAT TAGGAGCTTC CTTACTGCTC
1101 CTCTGGTGA TCCAAGTGTG GTGGGACCCC TACTAGGGT CAGGAAGTGG
1151 AACTAACAT CTGTGAGGT GTTGACTTGA AAAATAAAGT GTTGATTGGC
1201 TAAAAAATAA AAAAAAATAA AAAAAAATAA AAGGGCGGCC GCTCTAGAGG
1251 ATCCAAAGCTT ACGTAAAAAA AAAAAAATAA AAG
```

## BLAST Results

No BLAST result

## Medline entries

No Medline entry

## Peptide information for frame 2

ORF from 461 bp to 814 bp; peptide length: 118  
Category: putative protein

```
1 MVAIPPSACL PACCPGHGAV PVPRIGFKFV NNFPFGLVDV NRAREVLPTA
51 CACLPASSLF SFHYAPSPGG LALSFSSYPQ GPVLLCPHVP LGCLVEALYN
101 FSLVLCFLL YFPAVSCP
```



## BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphtes3\_15c6, frame 2

PIR:S54250 ribosomal protein L2 - Arabidopsis thaliana, N = 1, Score = 76, P = 0.33

>PIR:S54250 ribosomal protein L2 - Arabidopsis thaliana  
Length = 258

## HSPs:

Score = 76 (11.4 bits), Expect = 4.0e-01, P = 3.3e-01  
Identities = 30/91 (32%), Positives = 44/91 (48%)

Query: 15 PGHGAVPVPRIGFKFVNFPFGLVDVNRAREVLTACACLPASSLFSFHYAPSPGGLALS 74  
PG GA P+ R+ F+ PF + +E+ A C P SSL+ A G L  
Sbjct: 52 PGRGA-PLARVTFRH----PFRF---KKQKELFVAAEVCTPVSSLYCGKKATLVVGNVLP 103

Query: 75 FSSYPQGPVLLCP---HV-PLGCLVEALYNFSLVL 105  
S P+G V+ C HV G L A ++++V+  
Sbjct: 104 LRSIFEGAVV-CNVEHHVGDGRGLARASGDYAIIV 137

Pedant information for DKFZphtes3\_15c6, frame 2

## Report for DKFZphtes3\_15c6.2

[LENGTH] 118  
[MW] 12413.79  
[pI] 7.53  
[PROSITE] LEUCINE\_ZIPPER 1  
[PROSITE] MYRISTYL 1  
[PROSITE] ASN\_GLYCOSYLATION 1  
[KW] TRANSMEMBRANE 1

SEQ MVAIPPSACLPACCPGHGAVPVPRIGFKFVNFPFGLVDVNRAREVLTACACLPASSLF  
PRD ccc  
MEM .....MM

SEQ SFHYAPSPGGLALSFSSYPQGPVLLCPHVPLGCLVEALYNFSLVLCSEFLLYFPAVSCP  
PRD eeeeeeeeeeeeeeeeeeeeeeeeeeeeecccccccccccccccccccccccccccccccccccc  
MEM .....MM

## Prosites for DKFZphtes3\_15c6.2

PS00001	100->104	ASN_GLYCOSYLATION	PDOC00001
PS00008	70->76	MYRISTYL	PDOC00008
PS00029	84->106	LEUCINE_ZIPPER	PDOC00029

(No Pfam data available for DKFZphtes3\_15c6.2)

DKFZphtes3\_15g14

-----

group: testes derived

DKFZphtes3\_15g14 encodes a novel 701 amino acid protein with weak similarity to *S. cerevisiae* hypothetical protein YOR243c.  
No informative BLAST results; No predictive prosite, pfam or SCOP motife.

The new protein can find application in studying the expression profile of testis-specific genes.

similarity to YOR243c

complete cDNA, complete cds, potential start codon at Bp 35, EST hits

Sequenced by GBF

Locus: unknown

Insert length: 3495 bp

Poly A stretch at pos. 3462, no polyadenylation signal found

```
1 GCCTTCCACT GAACCGAGGC ACTGTTATAG AAGAATGGAA GAAGATACAG
51 ATTATAGAAT CAGGTTTAGT TCTTTGTGTT TCTTTAATGA TCACGTTGGA
101 TTTTCATGGCA CTATAAAAAG CTCACCAAGT GACTTTATGT TTATTGAAAT
151 TGATGAACAG GGACAGTTAG TTAATAAGAC CATCGATGAG CCTATTTTCA
201 AGATTAGTGA AATACAACCT GAGCCAAATA ATTTCCCAA AAAACCAAAA
251 CTAGATCTTC AAAATCTGTC CTAGAAGAT GGAAGAAACC AAGAAGTTCA
301 TACTTTGATT AAGTACACTG ATGGTGACCA AAATCATCAG TCTGGTTCAG
351 AAAAGGAAGA TACTATCGTT GATGGAACTT CCAATGTGA AGAAAAAGCT
401 GATGTTTTAA GCTCCTTTT GGATGAAAAA ACTCATGAGT TACTGAATAA
451 TTTTGCCTGT GATGTAAGAG AGAAGTGGCT TTCTAAAAA GAGCTAATTG
501 GACTACCTCC TGAATTCTCA ATAGGCAGAA TCCTTGACAA AAACCAGAGG
551 GCTAGTTTAC ACAGTGCCAT TAGGCAGAAA TTTCCATTTT TAGTAACGTG
601 AGGAAAAAAC AGTGAATTTG TTGTAAACC AAATCTTGAA TATAAAGAAC
651 TTTGTCAATT GGTATCTGAA GAGGAAGCAT TTGACTTTT TAAATATTTG
701 GATGCAAGA AAGAAAAATC CAAATTTACC TTTAAACCTG ATACAAACAA
751 AGACCACAGA AAGCTGTGCC ACCATTTTGT CACAAAAAG TTTGGAACCC
801 TTGTGGAAC CAAATCTTTT TCTAAAATGA ATTGCAGTGC TGGTAATCCG
851 AATGTGGTGG TAACAGTAAG ATTTCCGGGA AAAGCACACA AACGTGGGAA
901 AAGGCCCTCT TCTGAATGCC AAGAAGGAAA AGTTATATAT ACAGCTTTTA
951 CCCTACGAAA GGAAAAACCTG GAAATGTTTG AAGCGATTGG TTTTGTAGCT
1001 ATCAAACTTG GTGTTATTCC TTCCGATTTT AGTTATGCAG GCCTTAAAGA
1051 CAAGAAAGCC ATCACTATC AAGCAATGGT TGTAGAAAA GTGACTCCAG
1101 AGAGGTTGAA AATATTGAA AAGAAATTG AAAAGAAAG AATGAATGTC
1151 TTTAATATTC GGTCTGTAGA TGATTCCTGT AGACTTGGTC AGCTCAAAAGG
1201 AATCACTTT GATATTGTCA TTAGAAATTT AAAAAACAA ATAAATGATT
1251 CTGCAACCT GAGGAGAGA ATTATGGAAG CAATAGAAAA TGTTAAGAAA
1301 AAAGGCTTTG TGAATTACTA TGGACCACAG AGATTGGGA AGGGAAGGAA
1351 AGTTCACACA GACCAAAATT GACTAGCTTT GCTGAAGAAT GAAATGATGA
1401 AAGCCATAAA ATTGTTTCTT ACACCAGAG ACTTGGATGA TCCTGTAAT
1451 AGAGCAAGA AGTATTTCTT TCAAACTGAG GATGCTAAAG GCACACTTTC
1501 ATTGATGCTT GAATTCAAAG TGGCTGAGAG AGCATTTGTT GAGGCATTGC
1551 ACCGCTTTGG CATGACCGAG GAAGTTGTTA TCCAGGCATG GTTCTCTTTA
1601 CCCCATTCCA TGCGCATATT CTATGTTTAC GCATATACCA GCAAAATTG
1651 GAATGAGGCA GTATCTTACA GACTTGAAAC CTATGGAGCA AGAGTAGTGC
1701 AGGGTGATTT GGTCTGTTTG GATGAAGACA TTGATGACGA GAATTTCCCA
1751 AATAGTAAAA TTCACTGGT AACTGAAGAG GAGGATCAG CTAATATGTA
1801 TGCAATACAT CAGGTGGTTC TTCCAGTACT TGGATACAAT ATTCACTACC
1851 CGAAGAACAA AGTAGGGCAG TGGTACCATG ACATACTTAG CAGAGATGGA
1901 CTACAGACAT GTAGGTTTAA AGTACCTACT CTGAACTGA ATATACCAGG
1951 TTGCTATAGA CAGATTTTGA AACATCCCTG TAATCTCTCA TACCAACTAA
2001 TGAAGATCA TGACATTGAT GTCAAAACGA AAGGTTCCCA CATTGATGAA
2051 ACAGCTTTGT CTCTTTTGT CTCTTTTGT CTGATGCTT CATGCTATGC
2101 TACCGTTTGT CTGAAGGAAA TAATGAAGCA TGACGTTTAA AACTGATACC
2151 CTTGGTATAA CCATATATAT GTCAACCTTT CCTGTTTTG AAATATTGTA
2201 TCAGAACAA ATACAAGGGA AATGCCATAC CTCTGTTGT GATAGATACC
2251 CCAGAGTAGT TATTACCTCT TTGTGAGATA AGTAATCTTT GATGAAGATT
2301 GAAATACAAT TTCTCATCCA ATTTTATAT CTTGGCATAC GCTGACCTTC
2351 TTGACCAATT GTAATTTTTT CATATTAAT AAAACAGGTG TTAGAGTCAG
2401 ACAGATTCAAT TCTTAGATTG TAGCTCTGAC ACTTACTAGT GATTTTGAGT
2451 ATGTTGTTGA TTTTTTTGTG TGTGGTTACT GATAGAATCA AGACAATTAC
2501 AACTTCATAA ATGACAAATA ATAGGATTAT CTCACATTT TCTGTTGCTG
2551 GAGGAACAAA ACATTGTGCC CATTGTAAAA TTTAATTTT TGTGTTGTTA
2601 ACTATCCAC ATTATAATC ATCCTTACC ATTTATATC AGTTAAATAT
2651 GGGTGTGTT GGGAGGAATG ACTGGCATG AGACATGTAT TGATTAGGA
2701 AGATCTGAGC ATTTCTTTCA TTGTTGGTAA GATATAATGA TGAATTTAA
```

```

2751 AAAGCAGTAT GGAGCATTAT ATATCAGTAA TGTGATATAT ATACTTAAGC
2801 CAGTTTAAAC ATTTTGGGAA ATGTTAGCAT TAGGAAATAA AATCCAAAAG
2851 AAGGAAGAGA AGCTATATGC AATGCAAAAT TTGCTTATTG CAATATTTTC
2901 ATATACAGAC ACTAAAACAA GTTTTCAAGG TCCAGCATTG CGTAACATAA
2951 GTAAGTAAAA TGATGTGTAT CAACCTGATG GTAAAATATG TAGTTATTTA
3001 AAAAAGCAAT GAACAATTTA GTTTCATGAG AAAATGTTGC CCCTTAAAAG
3051 TAGAACACAT ATGTTACAAC TGCAATAATA CTCTGAATTC ATCTTTCACA
3101 AATAAGAGAC ATGTTAGCAT AGTGATTAAT AGCAGAGATA TTGAGAGCAA
3151 ACTAACCCAG TTGAAACCCG GGCACCTGCC CGTATAGCAC TGCAGCCTTG
3201 GGAAAGTTAT TTAACCTCAT GGGCTTCAGT TTCAACATCT GTAAAATGGG
3251 CATGTTAACA TTGCCTACCT CATAGGATTA CTGTGAGAAAT TTTCTAAGTT
3301 AATATATGTA AAGCAACTTT AAAAAGTGCC TGGCACTTAG TTATTGTTAA
3351 GTAAGTGTCT GCAGATGCAA GTTTGGAAGA GAAAAGCAA TAAATGAAAA
3401 TCCCTTCCTG TTAAGATGAA AAAAAAAAAA AAAAAAAAAA AAAAAAGGGG
3451 CGGCCGCTCA AGATGAAAAA AAAAAAAAAA AAAAAAAAAA AAAGG

```

## BLAST Results

No BLAST result

## Medline entries

No Medline entry

## Peptide information for frame 2

ORF from 35 bp to 2137 bp; peptide length: 701  
 Category: similarity to unknown protein

```

1 MEEDTDYRIR FSSLCFFNDH VGFHGTIKSS PSDFIVIEID EQQLVNKTI
51 DEPIFKISEI QLEPNFPKK PKLDLQNLSD EDGRNQEVT LIKYTDGQDN
101 HQSGSEKEDT IVDGTSKCEE KADVLSSFLD EKTHELLNNF ACDVREKWL
151 KTELIGLPPE FSIGRILOKN QRASLHSAIR QKFFFLVTVG KNSIVVKPN
201 LEYKELCHLV SEEEAFDFEK YLDAKKENSK FTFFKPDNKH HRKAVHHFVN
251 KKGFLNVETK SFSKMNCSAG NPNVVTVRF REKAHKGKR PLSECQEGKV
301 IYTAFTLRKE NLEMFPAIGF LAIKLGIVPS DFSYAGLKDK KAITYQAMV
351 RKVTPERLKN IEKEIEKKRM NVFNIRSVDD SLRLGQLKGN HFDIVIRNLK
401 QINDSANLR ERIMEAIEHV KKGFFVNYG PQRFGKGRKV HTDQIGLALL
451 KNEMMKAIKL FLTPEDLDDP VNRAKYFLQ TEDAKGTLSL MPEFKVRERA
501 LLEALHFRGM TEEGCIQAMF SLPHSMRIFY VHAYTSKIWN EAVSYRLETY
551 GARVVQGDV CLDEDIDDEN FPNKIHVLT EEEGSANMYA IHQVLPVLG
601 YNIQYQKNKV GQWYHDILSR DGLQTCRFKV PTLKLNIPGC YRQILKHPCN
651 LSYQLMEDHD IDVKTGSHI DETALSLIIS FDLBASCYAT VCLKEIMKHD
701 V

```

## BLASTP hits

No BLASTP hits available

## Alert BLASTP hits for DKF2phtes3\_15g14, frame 2

TREMBL:SPBC1A45P\_10 gene: "SPBC1A4.09"; product: "hypothetical protein"; S.pombe chromosome II cosmid clA4 left hand region 1-26184 bp  
 Originates from chimeric cosmid., N = 3, Score = 511, P = 2.9e-57

PIR:S67136 hypothetical protein YOR243c - yeast (Saccharomyces cerevisiae), N = 2, Score = 516, P = 7.3e-54

SWISSPROT:YQ4B\_CAEEL HYPOTHETICAL 64.6 KD PROTEIN B0024.11 IN CHROMOSOME V., N = 2, Score = 386, P = 2.1e-34

>PIR:S67136 hypothetical protein YOR243c - yeast (Saccharomyces cerevisiae)  
 Length = 676

## HSPs:

Score = 516 (77.4 bits), Expect = 7.3e-54, Sum P(2) = 7.3e-54  
 Identities = 151/498 (30%), Positives = 245/498 (49%)

Query: 191 KNSIVVKPNLEYKELCHLVSEEEAFDFEK-YLDAKKENSKFTFFKPDNKHHRKAVHHFV 249  
 + E V P L +L + EE+ Y A K + F+ +K R +H +

Sbjct: 109 RRQEFNVDPRLR-NQLVEIFGEEDVLKIESVYRTANKMETAKNFE---DKSVRTKIHQLL 164

Query: 250 NKKFGNLVETKSFSSKMNCAGNPNVVTVRFREKAHK-RGKRPLSECQEG-KVIYTAFTL 307  
+ F N +E+ + N +EK ++ R + G + FTL

Sbjct: 165 REAFKNELESVTTDTNTFKIARSNRNRSTNKQEKINQTRDANGVENWGYGPKDFIHTL 224

Query: 308 RKENLEMFEAIGFLAIKLGVIPSD-FSYAGLKDKKAITQAMVVRKVTPERLKNIKEIE 366  
KEN + EA+ + KL +PS YAG KD++A+T Q + + K+ +RL + + +

Sbjct: 225 HKENKDTMEAVNVIT-KLLRVPSRVIRYAGTKDRRAVTCQRVISIKIGLDRNLALNRTL- 282

Query: 367 KKRNVFNIRSVDDSLRLGQLKGNHFDIVIRNLKKQINDSANLRERIMEAIENVKKKGFV 426  
K M + N D SL LG LKGN F +VIR++ N +L E + +++ + GF+

Sbjct: 283 -KGMIGNYFNFSASLNLGDLKGNFVVVIRDVTTG-NSEVSLEEIVSNGCKSLSENGFI 340

Query: 427 NYYGQRFQGRKRVHTDQIGLALLKNEMKAIKFLTPEDLODPVNR-AKKYFLOTEDAK 485  
NY+G QRFQ + T IG LL + KA +L L+ +D P ++ A+K + +T+DA

Sbjct: 341 NYFGMQRFQGT-FSISTHTIGRELLSNWKKAAELILSDQDNVLPKSKEARKIWAETKDAA 399

Query: 486 GTLSLMPEFKVRERALLEALHFRFGMTEEGCIQ--AWFS----LPHSMRIFYVHAYTSKI 539  
L MP + E ALL +L E+G A+++ +P ++R YVHAY S +W

Sbjct: 400 LALKQMPROCLAENALLYSLSNQRKEEDGTYSENAYTAIMKIPRNLRTMYVHAYQSYVW 459

Query: 540 NEAVSYRLETYGARVVQGDVLC-----LDEDIDDENFPNS-----KIHVTEEEGS 585  
N S R+E +G ++V GDLV L IDDE+F + VT+E+

Sbjct: 460 NSIASKRIELHGLKLVVGDVLDITSEKSLISGIDDEDFOEDVREAQFIRAKAVTQEDID 519

Query: 586 ANMYAIHQVLPVLGYNIQYPKNK-VGQWYHDILSRDGLQTCRFKVPKLKLNIPGCYRQI 644  
+ Y + VVLP G+++ YP N+ + Q Y DIL D + + ++ G YR +

Sbjct: 520 SVKYTMEDVVLFPSPGFDVLYPSNEELKQLYVDILKADNMDFNMRKVRDFSLAGSYRTV 579

Query: 645 LKHPCNLSYQLMEDHDIDVKTGSHID 671  
++ P +L Y+++ D + + + +D

Sbjct: 580 IQPKSLLEYRIIHYDDPSQQLVNTDLD 606

Score = 86 (12.9 bits), Expect = 3.2e-01, Sum P(2) = 2.8e-01  
Identities = 40/160 (25%), Positives = 77/160 (48%)

Query: 22 GFHGTIKSSPSDFIVIEIDEQQQLVNKTIDEPIFKISEIQLEPNFPPKPKLDQLNLSLE 81  
GF G IK +DF+V EID++G++++ T D+ FK+ + +P K +++ + S E

Sbjct: 55 GFRGQIKQRYTDFLVNEIDQEGKVIHLT-DKG-FKMPK---KPQR--SKEEVNAEKES-E 106

Query: 82 DGRNQEVHTLIKTYDGDQNHQSGS--EKEDTI-VDGTSKCEEKADVLSSFLDEKTHELLN 138  
R QE + D + +Q +ED + ++ + K + +F D+ ++

Sbjct: 107 AARRQEFNV----DPELRNQLVEIFGEEDVLKIESVYRTANKMETAKNFEDKSVRTKI 161

Query: 139 NFACDVREKWLKTELGILPPE-FSIGRILDKNQASLHSAIRQ 181  
+RE + ++ E + F I R ++N R + I Q

Sbjct: 162 QL---LREAFKNELESVTTDTNTFKIARS-NRNRSTNKQEKINQ 201

Score = 58 (8.7 bits), Expect = 7.3e-54, Sum P(2) = 7.3e-54  
Identities = 10/23 (43%), Positives = 17/23 (73%)

Query: 676 SLLISFDLDASCYATVCLKEIMK 698  
++++ F L S YAT+ L+E+MK

Sbjct: 638 AVVLKFLGLTSAYATMALRELK 660

Pedant information for DKF2phtes3\_15g14, frame 2

Report for DKF2phtes3\_15g14.2

{LENGTH} 701  
{MW} 80700.96  
{PI} 7.31  
{HOMOL} PIR:S67136 hypothetical protein YOR243c - yeast (Saccharomyces cerevisiae) 2e-51  
{FUNCAT} 99 unclassified proteins [S. cerevisiae, YOR243c] 8e-53  
{BLOCKS} BL01268C  
{BLOCKS} BL01268B  
{BLOCKS} BL01268A  
{SUPFAM} hypothetical protein HI0701 3e-06  
{PROSITE} MYRISTYL 7  
{PROSITE} AMIDATION 2  
{PROSITE} CAMP\_PHOSPHO\_SITE 1  
{PROSITE} CK2\_PHOSPHO\_SITE 16  
{PROSITE} TYR\_PHOSPHO\_SITE 1  
{PROSITE} PKC\_PHOSPHO\_SITE 13  
{PROSITE} ASN\_GLYCOSYLATION 5  
{KW} Alpha\_Beta

```

SEQ MEEDTDYRIRFSSLCFFNDHVGFGHTIKSSPSDFIVIEIDEQGLVNKTIDEPIFKISEI
PRD ccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc

SEQ QLEPNFPKKPKLDLQNLSELDGRNQEVHTLIKYTDGDNHQSGSEKEDTIVDGTSKCEE
PRD ccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc

SEQ KADVLSSFLDEKTHELLNNFACDVREKWLKTELGPPFESIGRILDKNQASLHSAIR
PRD hhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh

SEQ QKFPFLVTVGKNSEIVVKPNLEYKELCHLVSEEAFFFKYLDAKKENSÁFTFKPDTNKG
PRD hhcccccccccccccccccccccccccccccccccccccccccccccccccccccccccc

SEQ HRKAVHHFVNKKFGNLVETKSF SKMNC SAGNPVVTVRFREKAHKGKRLPSECQEGKV
PRD hhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh

SEQ IYTAFTLRKENLEMFAGFLAIGLGVIPSDFSYAGLKDKKAITYQAMVVRKVTPERLKN
PRD eeeeecccccccccccccccccccccccccccccccccccccccccccccccccccccc

SEQ IEKEIEKRMNVFNIRSVDDSLRLGQLKGNHFDIVIRNLKKQINDSANLRERIMEAENV
PRD hhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh

SEQ KKGGFVNYGQRFGRGKRVHTDQIGLALLKNEMMKAIKLFLTPEDLDDPNRAKKYFLQ
PRD hhcccccccccccccccccccccccccccccccccccccccccccccccccccccccccc

SEQ TEDAKGTLMLPEFKVRERALLEALHRFGMTEEGCIQAWFSLPHSMRI FYVHAYTSKIWN
PRD hcccccccccccccccccccccccccccccccccccccccccccccccccccccccccc

SEQ EAVSYRLETYGARVVQDGLVCLDEDIDENFPNSKIHLVTEEGSANMYAIHQVVLVPLG
PRD hhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh

SEQ YNIQYPKNKGQWYHDI LSRDGLQTCRFKVP TLKLNIPGCYRQILKHPCLNLYQLMEDHD
PRD ccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc

SEQ IDVKTGSHIDE TALSLLISFDLASCYATVCLKEIMKHVD
PRD ccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc

```

## Prosite for DKFZphtes3\_15g14.2

PS00001	47->51	ASN_GLYCOSYLATION	PDOC00001
PS00001	77->81	ASN_GLYCOSYLATION	PDOC00001
PS00001	266->270	ASN_GLYCOSYLATION	PDOC00001
PS00001	404->408	ASN_GLYCOSYLATION	PDOC00001
PS00001	650->654	ASN_GLYCOSYLATION	PDOC00001
PS00004	351->355	CAMP_PHOSPHO_SITE	PDOC00004
PS00005	26->29	PKC_PHOSPHO_SITE	PDOC00005
PS00005	105->108	PKC_PHOSPHO_SITE	PDOC00005
PS00005	115->118	PKC_PHOSPHO_SITE	PDOC00005
PS00005	232->235	PKC_PHOSPHO_SITE	PDOC00005
PS00005	237->240	PKC_PHOSPHO_SITE	PDOC00005
PS00005	277->280	PKC_PHOSPHO_SITE	PDOC00005
PS00005	306->309	PKC_PHOSPHO_SITE	PDOC00005
PS00005	381->384	PKC_PHOSPHO_SITE	PDOC00005
PS00005	525->528	PKC_PHOSPHO_SITE	PDOC00005
PS00005	535->538	PKC_PHOSPHO_SITE	PDOC00005
PS00005	544->547	PKC_PHOSPHO_SITE	PDOC00005
PS00005	625->628	PKC_PHOSPHO_SITE	PDOC00005
PS00005	632->635	PKC_PHOSPHO_SITE	PDOC00005
PS00006	30->34	CK2_PHOSPHO_SITE	PDOC00006
PS00006	49->53	CK2_PHOSPHO_SITE	PDOC00006
PS00006	79->83	CK2_PHOSPHO_SITE	PDOC00006
PS00006	95->99	CK2_PHOSPHO_SITE	PDOC00006
PS00006	103->107	CK2_PHOSPHO_SITE	PDOC00006
PS00006	105->109	CK2_PHOSPHO_SITE	PDOC00006
PS00006	110->114	CK2_PHOSPHO_SITE	PDOC00006
PS00006	116->120	CK2_PHOSPHO_SITE	PDOC00006
PS00006	127->131	CK2_PHOSPHO_SITE	PDOC00006
PS00006	150->154	CK2_PHOSPHO_SITE	PDOC00006
PS00006	211->215	CK2_PHOSPHO_SITE	PDOC00006
PS00006	237->241	CK2_PHOSPHO_SITE	PDOC00006
PS00006	377->381	CK2_PHOSPHO_SITE	PDOC00006
PS00006	463->467	CK2_PHOSPHO_SITE	PDOC00006
PS00006	580->584	CK2_PHOSPHO_SITE	PDOC00006
PS00006	668->672	CK2_PHOSPHO_SITE	PDOC00006
PS00007	537->546	TYR_PHOSPHO_SITE	PDOC00007
PS00008	25->31	MYRISTYL	PDOC00008
PS00008	43->49	MYRISTYL	PDOC00008
PS00008	114->120	MYRISTYL	PDOC00008

WO 01/12659

PCT/IB00/01496

PS00008	326->332	MYRISTYL	PDOC00008
PS00008	385->391	MYRISTYL	PDOC00008
PS00008	514->520	MYRISTYL	PDOC00008
PS00008	622->628	MYRISTYL	PDOC00008
PS00009	287->291	AMIDATION	PDOC00009
PS00009	436->440	AMIDATION	PDOC00009

(No Pfam data available for DKFZphtes3\_15g14.2)

DKFZphtes3\_15h1

-----

group: testes derived

DKFZphtes3\_15h1 encodes a novel 672 amino acid protein with very weak similarity to several proteins.

No informative BLAST results; No predictive prosite, pfam or SCOP motif.

The new protein can find application in studying the expression profile of testis-specific genes.

similarity to Hsp70/Hsp90 organizing protein

complete cDNA, complete cds, no EST hits

Sequenced by GBF

Locus: unknown

Insert length: 2277 bp

Poly A stretch at pos. 2252, polyadenylation signal at pos. 2226

```
1 AAACAGATA GAGGTTCTCC AGCTTTTCTT TGATTGTCTC TGCTTTAGCG
51 TCTCTAAATC CGGTACCAT GTCCGACCCC GAAGGCGAGA CCTTGCAGAG
101 CACCTTTCCC TCTTATATGG CCGAAGGCGA GCGGCTCTAC CTGTGCGGGG
151 AATTTTCTAA AGCCGCGCAG AGCTTCAGCA ACGCTCTTTA CCTTCAGGAT
201 GGAGACAAGA ACTGCCTGGT TGCTCGCTCA AAGTGCTTCC TGAAGATGGG
251 AGACTTGGAG AGATCCCTGA AGGATGCTGA GGCTTCGCTC CAGAGTGACC
301 CAGCTTTCTG TAAGGGGATT TTGCAAAAGG CTGAGACACT GTACACCATG
351 GGAGACTTTG AGTTTGCCTT GGTATTCTAT CATCGAGGCT ACAAGCTGAG
401 GCCTGATCGG GAATTCAGAG TTGGCATTCA GAAAGCCCAG GAAAGCATCA
451 ACAACTCAGT GGAAGTCCT TCTTCCATTA AGCTGGAGAA CAAAGGGGAC
501 CTCTCCTTCT TAAGCAAGCA GGCTGAGAA ATAAAAGCCC AGCAGAAGCC
551 TCAGCCCATG AACACCTCT TACACCCAC CAAGGGAGAG CCCAAGTGA
601 AGGCCTCGCT CAAGAGTGAG AAGACTGTCC GCCAGCTTCT GGGGAGCTC
651 TACGTGGACA AAGAGTATT GGAGAAGCTC CTATTGGATG AAGACCTGAT
701 CAAAGGCACC ATGAAGGGCG GCCTGACTGT GGAGGACCTC ATCATGACCG
751 GCATCAACTA CTTGGATACT CACAGCAACT TCTGGAGGCA GCAGAAGCCG
801 ATCTACGCCA GGGAGCGGGA CCGGAAGCTG ATGCAAGAGA AATGGCTCGG
851 GGACCACAAA CGCCGTCCCT CACAGACAGC CCATTACATC CTCAGAGCC
901 TGGAGGACAT TGATATGTTG CTCACAAGTG GCAGTGCTGA AGGGAGTCTT
951 CAGAAAAGCTG AGAAAAGTCT GAAGAAGTGA CTGGAATGGA ACAAGGAAGA
1001 GGTACCCAAC AAGGATGAAC TGGTTGGAAG CTGTATAGC TGCAATAGGA
1051 ATGCCCAGAT TGAGCTGGGG CAGATGGAGG CAGCCCTGCA GAGCCACAGA
1101 AAGGACCTGG AGATCGCCAA GGAATATGAC CTTCTGATG CAAAATCGAG
1151 AGCCCTTGAC AACATTGGCA GAGTTTTTGC CAGAGTTGGG AAATTCAGC
1201 AAGCCATTGA CACGTGGGAA GAAAAGATCC CTCTGGCAAA AACCACCTG
1251 GAGAAAGCCT GGCTGTTCCA CGAGATCGGC CGCTGCTACT TGGAGCTGGA
1301 CCAGGCCTGG CAGGCCCAGA ATTATGGCGA GAAGTCCAG CAGTGTGCCG
1351 AGGAGGAAGG GGACATTGAG TGGCAACTGA ATGCCAGTGT TCTGGTGGCC
1401 CAGGCACAAG TGAAGCTGAG AGACTTCGAG TCAGCCGTGA ACAATTTTGA
1451 GAAGGCCCTG GAGAGAGCAA AGCTTGTGCA TAACAACGAG GCGCAGCAGG
1501 CCATCATCAG TGCCTTGGCA GATGCCAACA AGGTATCAT CAGAGAAGCTG
1551 AGGAAAACCA ACTACGTGGA GAATCTCAA GAAAAAGCG AGGGAGAAGC
1601 TTCACTGTAT GAAGATAGAA TAATAACAAG AGAGAAGGAC ATGAGGAGAG
1651 TGAGAGATGA GCCCAGAGA GTGTTGAAGC AGTGGGACCA TAGTGAGGAT
1701 GAGAAAGAGA CAGATGAGGA CGATGAGGCT TTTGGGGAAG CTCTGCAGAG
1751 CCCAGCAAGC GGAAGCAGA GTGTGGAAGC AGGAAAAGCC AGAAGCGATT
1801 TGGGAGCAGT TGCCAGGGC CTGTACGAGG AATTAGGCAC AAGATCAGGA
1851 GAAACAGGCA GGAAGCTACT AGAAGCTGGC AGAAGAGAGT CAAGAGAAAT
1901 TTATAGGAGG CTTTCGGGAG AATTAGAGCA AAGACTCTCA GGAGAATTCA
1951 GCAGACAGGA ACCAGAAGAA CTAAAG.AAC TTTCAGAAAT GGGCAGAAGA
2001 GAGCCAGAAG AACTGGGAAA AACACAATTT GGAGAAATAG GAGAAACGAA
2051 AAAAAACAGGA AATGAGATGG AAAAGGAATA TGAATGAAGC CATCGGTAGA
2101 GATGAGGATC AGGAAGCTGG TGTTCAGAGG GATCATGGGA TTTTATTAAA
2151 CTGGATTTTC AAGCGATTGG TCTGTTATAG GAAAAATGAG GGTTTTACTT
2201 CTGCTGCTTT CCATCACTAT TTTGCCATTA AATAGGTGTC TTTCACCTTT
2251 GCAAAAAAAA AAAAAAAA AAAAAA
```

## BLAST Results

-----

No BLAST result

## Medline entries

No Medline entry

## Peptide information for frame 3

ORF from 69 bp to 2084 bp; peptide length: 672  
Category: similarity to known protein

1 MSDPEGETLR STFPSYMAEG ERLYLCEFS KAAQSFSNAL YLQDGDKNCL  
51 VARSKCFMK GDLERSLKDA EASLQSDPAF CKGILQKAET LVTMGDFEFA  
101 LVFYHRGYKL RPDREFRVGI QKAQEAINNVS VGSPSSIKLE NKGDLSFLSK  
151 QAENIKAAQK PQPMKHLHP TKGEPRWKAS LKSEKTVRQL LGELYVDKEY  
201 LEKLLDDEL IKGTMKGGIT VEDLIMTGIN YLDTHSNFWR QQKPIYARER  
251 DRKLMQEKWL RDHKRRPSQT AHYILKSLED IDMLTSGSA EGSQKAEKV  
301 LKKVLEWNKE EVPNKDELVG NLYSCIGNAQ IELQMEAAAL QSHRKDLTA  
351 KEYDLPDAKS RALDNIGRVF ARVGKFFQAI DTWEEKIPLA KTTLEKTWLF  
401 HEIGRCYLEL DQAWQAQNYG EKSQQCAEEE GDIEWQLNAS VLVQAQVVKL  
451 RDFSAVNNF EKALERAQKV HNNEAQAII SALDDANKGI IRELKRTNYV  
501 ENLKEKSEGE ASLYEDRIIT REKDMRRVRD EPEKVVQWD HSEDEKETDE  
551 DDEAFGEALQ SPASGKQSV EAGKARSDLGA VAKGLSGELG TRSGETGRKL  
601 LEAGRRESRE IYRRPSGELE QRLSGEFSRQ EPEELKKLSE VGRREPEELG  
651 KTQFGIEGT KKTGNEMEKE YE

## BLASTP hits

Entry AF039202.1 from database TREMBL:  
product: "Hsp70/Hsp90 organizing protein"; Cricetulus griseus  
Hsp70/Hsp90 organizing protein mRNA, complete cds.  
Score = 149, P = 5.3e-07, identities = 42/160, positives = 74/160

Entry AI09782.1 from database TREMBL:  
product: "myosin heavy chain"; Argopecten irradians myosin heavy chain  
mRNA, complete cds.  
Score = 155, P = 6.1e-07, identities = 140/623, positives = 256/623

Entry S56658 from database PIR:  
stress-induced protein stil - soybean  
Score = 156, P = 9.7e-08, identities = 41/153, positives = 72/153

## Alert BLASTP hits for DKFZphtes3\_15h1, frame 3

No Alert BLASTP hits found

## Pedant information for DKFZphtes3\_15h1, frame 3

## Report for DKFZphtes3\_15h1.3

[LENGTH] 672  
[MW] 76655.61  
[pI] 5.49  
[HOMOL] PIR:S56658 stress-induced protein stil - soybean 6e-10  
[SUPFAM] tetratricopeptide repeat homology 1e-07  
[PROSITE] MYRISTYL 7  
[PROSITE] AMIDATION 3  
[PROSITE] CAMP\_PHOSPHO\_SITE 4  
[PROSITE] CK2\_PHOSPHO\_SITE 15  
[PROSITE] TYR\_PHOSPHO\_SITE 1  
[PROSITE] PKC\_PHOSPHO\_SITE 11  
[PROSITE] ASN\_GLYCOSYLATION 2  
[KW] All\_Alpha  
[KW] LOW\_COMPLEXITY 4.76 %

SEQ MSDPEGETLRSTFPSYMAEGERLYLCGEFSKAAQSFSNALYLQDGDKNCLVARSKCFMK  
SEG .....  
PRD cccccccceccccccccccccccccchhhhhhhhhhhhhccccceehhhhhhhhh  
  
SEQ GDLERSLKDAEASLQSDPAFCGILQKAETLYTMGDFEFALVFYHRGYKL RPDREFRVGI  
SEG .....  
PRD hcchhhhhhhhhhhhhccchhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhccccchhhhh



```

SEQ SEG PRD QKAQAEAINNSVSGSPSSI KLEKNKGDL SFLSKQAENI KAAQQPOPMKHLHPTKGFPKWKAS
SEQ SEG PRD hhhhhhhhhhhhhhhhhhhhhhhccchhhhhhhchhhhhhhhhccchhhhhhhccccccccchhh
SEQ SEG PRD LKSEKTVRQLGELYVDKEYLEKLLDLEDIKGTMKGGLTVEDLIMTGINYL DTHSNFWR
SEQ SEG PRD . . . xxxxxxxxxxxxxxxxxxxx
SEQ SEG PRD hhhhhhhhhhhhhhhhhhhhhhhccchhhhhhhhhhhhhhhhhhhhhcccccccccccccc
SEQ SEG PRD QQKPIYARERDRKLMQEKWL RDHKRRPSQTAYHILKSLEDIDMLTSGSAEGSLQAKAEV
SEQ SEG PRD cchhhhhhhhhhhhhhhhhhhhhccccccchhhhhhhhhhhhhheeeccccchhhhhhhhh
SEQ SEG PRD LKKVLEWNKEEVPNKDEL VGNLYSCIGNAQI ELQGEAAALQSHRKDLEIAKEYDLPKAS
SEQ SEG PRD hhhhhhhhhccccccccceccccccccchhhhhhhhhhhhhhhhhhhhhhhhhhhccccch
SEQ SEG PRD RALDNI GRVFA RVGK FQQAIDTWE EKIPLAKTTLEKTWLFHEIGRCYELDQAWQAQNYG
SEQ SEG PRD hhhccchhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhccchhhhhhhhhhhhh
SEQ SEG PRD EKSQQCAEEGGDI EWQLNASVLVAQAVKLRDFESA VNNFEKALERAKLVHNNEAQAQAI
SEQ SEG PRD hhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhccchhhhh
SEQ SEG PRD SALDDANKGII RELRKTNYVENL KEKSEGEASLYEDRIITREKDMRRVRDEPEKVVKQMD
SEQ SEG PRD hhhccchhhhhhhhhhhhhhhhhhhhhhhccchhhhhhhhhhhhhhhhhhhhhcccccccc.X
SEQ SEG PRD HSEDEKTEDEDDAEFGALQSPASGQSV EAGKARS DLGAVAKGLSGELGTRSGETGRKL
SEQ SEG PRD xxxxxxxxxxxxxx
SEQ SEG PRD cccccccccchhhhhhhhhccccccccchhhhhccccccccceeeccccccccccccchhh
SEQ SEG PRD LEAGRRESRIYRPSGELEQRLSGFSGRQEP EELKKLSEVGRPREPEELGKTGFGEIGET
SEQ SEG PRD hhhccccccccceccccchhhhhccccccchhhhhhhhhhhhhcccccccccccccccccc
SEQ SEG PRD KKTGNMEKEYE
SEQ SEG PRD . . . . .
SEQ SEG PRD cccccccccccc

```

Prosite for DKFZphtes3\_15h1.3

PS000001	128->132	ASN_GLYCOSYLATION	PD0C00001
PS000001	438->442	ASN_GLYCOSYLATION	PD0C00001
PS000004	265->269	CAMP_PHOSPHO_SITE	PD0C00004
PS000004	605->609	CAMP_PHOSPHO_SITE	PD0C00004
PS000004	613->617	CAMP_PHOSPHO_SITE	PD0C00004
PS000004	636->640	CAMP_PHOSPHO_SITE	PD0C00004
PS000005	8->11	PKC_PHOSPHO_SITE	PD0C00005
PS000005	66->69	PKC_PHOSPHO_SITE	PD0C00005
PS000005	136->139	PKC_PHOSPHO_SITE	PD0C00005
PS000005	180->183	PKC_PHOSPHO_SITE	PD0C00005
PS000005	183->186	PKC_PHOSPHO_SITE	PD0C00005
PS000005	186->189	PKC_PHOSPHO_SITE	PD0C00005
PS000005	214->217	PKC_PHOSPHO_SITE	PD0C00005
PS000005	342->345	PKC_PHOSPHO_SITE	PD0C00005
PS000005	564->567	PKC_PHOSPHO_SITE	PD0C00005
PS000005	596->599	PKC_PHOSPHO_SITE	PD0C00005
PS000005	660->663	PKC_PHOSPHO_SITE	PD0C00005
PS000006	2->6	K2_PHOSPHO_SITE	PD0C00006
PS000006	66->70	K2_PHOSPHO_SITE	PD0C00006
PS000006	93->97	K2_PHOSPHO_SITE	PD0C00006
PS000006	171->175	K2_PHOSPHO_SITE	PD0C00006
PS000006	220->224	K2_PHOSPHO_SITE	PD0C00006
PS000006	277->281	K2_PHOSPHO_SITE	PD0C00006
PS000006	382->386	K2_PHOSPHO_SITE	PD0C00006
PS000006	392->396	K2_PHOSPHO_SITE	PD0C00006
PS000006	481->485	K2_PHOSPHO_SITE	PD0C00006
PS000006	507->511	K2_PHOSPHO_SITE	PD0C00006
PS000006	512->516	K2_PHOSPHO_SITE	PD0C00006
PS000006	542->546	K2_PHOSPHO_SITE	PD0C00006
PS000006	548->552	K2_PHOSPHO_SITE	PD0C00006
PS000006	628->632	K2_PHOSPHO_SITE	PD0C00006
PS000006	663->667	K2_PHOSPHO_SITE	PD0C00006
PS000007	506->515	TYR_PHOSPHO_SITE	PD0C00007
PS000008	119->125	MYRISTYL	PD0C00008
PS000008	132->138	MYRISTYL	PD0C00008
PS000008	213->219	MYRISTYL	PD0C00008

WO 01/12659

PCT/IB00/01496

PS00008	288->294	MYRISTYL	PDOC00008
PS00008	320->326	MYRISTYL	PDOC00008
PS00008	334->340	MYRISTYL	PDOC00008
PS00008	590->596	MYRISTYL	PDOC00008
PS00009	596->600	AMIDATION	PDOC00009
PS00009	603->607	AMIDATION	PDOC00009
PS00009	641->645	AMIDATION	PDOC00009

(No Pfam data available for DKFZphtes3\_15h1.3)

DKFZphtes3\_15i5  
-----

group: cell structure and motility

DKFZphtes3\_15i5 encodes a novel 717 amino acid protein with similarity to radial spokehead proteins.

The novel protein is similar to the *Chlamydomonas reinhardtii* radial spokehead protein of flagella or axoneme and to the *Strongylocentrotus purpuratus* sea urchin spermatozoa protein p63. This protein is important for the maintenance of a planar form of sperm flagellar beating. In addition, the novel protein contains a transferrin signature 1 for iron-binding. The new protein seems to be a part of the human radial spoke heads in spermatozoa.

BLAST results; No predictive prosite, pfam or SCOP motif.

The new protein can find application in modulating the structure of the human spermatozoa radial spoke head and modulation of sperm motility in men.

strong similarity to "radial spokehead" proteins

complete cDNA, complete cds, 1 EST hit (from a testis library)  
"radial spokehead" part of flagella in *Chlamydomonas*, this protein seems to be part of the sperm motor or tail

Sequenced by GBF

Locus: unknown

Insert length: 2478 bp

Poly A stretch at pos. 2452, polyadenylation signal at pos. 2433

```
1 CACCCCTGGCC CGCTCCCCGC GCCCTCCACG GGTAAACGGCC CCCTCTCTCG
51 GTGCTCAGAA ACCGGCCGGTG TCGACAGGTG GCTCTCGCTT GGCCTCCTTG
101 TCTGCAAGCC TTTCTCCTAG AGATCTGTGC CTCCTGGGGA ACCATGGGAG
151 ACCTGCGGCC CTACCTTGAG CGCCCTGCCC AGCAGCCTCC GGGCCGGAGG
201 ACTTCTCAGG CCTCCCAGAG GCGGCACAGT GGGGACCAAG CTCAGGCCCT
251 GGCAGCGGAC CCGAGGAGAG GGCAGCAGAT AGCTCCAGAC GCCACGCGAA
301 AGCCCTCTGG TTGGTCACAG AGGGGCGAGCC TGTCCTCAAC GGAGAACTTG
351 CTGATGCCCC AGGTCTTCCA GGCTGAGGAA GCGCGGCTGG CTGGCATGGA
401 GTACCCATCT GTGAACACGG GCTTTCCTTC AGAGTTCACG CTCAGCCCTT
451 ACTCTGATGA AAGCAGGATG CAGGTGCGCG CTGTTCACAG AACTGGACCC
501 CTGCAAGCGG TCCAGCAGGG CCAAGCAGC CTGTTCACAG AACTGGACCC
551 CACCTTCCAG GAGCCCCAG TCAACCCCTT AGGGCCCTTA CATAAGGAT
601 AGACAGACCA GTTCTCTGAA GGTGCCAGC AGGGCCCTTA CATAAGGAT
651 GACCCCTGCC TTCACTTCTT GCCCTCTGAG CTGGGCTTCC CACACTACAG
701 TGCCCAAGTG CCTGAGCCCG AGCCTCTGGA GCTGGCCGTG CAGAAGCCCA
751 AGCCCTACCT GCTGCAGACC AGCATCAATT GCGACCTCAG CCTGTACGAG
801 CACCTGGTAA ATCTGCTGAC CAAGATCCTG AACCAGCGCG CTGAGGACCC
851 CTGTCTGTGC CTGGAGTCTG TCAACCGCAC CACCGACTGG GACTGGTTC
901 ACCCCAGACT GGACACGCTG CCGGACGACC CCGAGATGCA GCCACCTAC
951 AAGATGGCGG AGAACAGAA GCGCTCTTTC ACCCGGATG GAGGCGGCAC
1001 TGAAGGCGAA CAGGACATGG AGGAGGAGGT GGGGGAGACA CCAATGCCCA
1051 ACATCATGGA GACTGCTTTC TACTTCGAGC AGGCCGGGCT CGGCTGAGC
1101 TCGGACGAGA GCTTCCGCAAT TTTCTGGGCC ATGAAACAGC TGGTGGAGCA
1151 GCAGCCCATC CACACCTCTC GCTTCTGGGG CAAGATCCTG GGAATCAAAC
1201 GCAGCTACTT GGTGCGCCAG GTGGAATTCC GGGAGGGGGA GGAGGAGGCA
1251 GAGGAGGAGG AGCTGCAGGA GATGACGGAA GGTGGCGAGG TCATGGAGGC
1301 GCACGGCGAG GAGGAGGGCG AGGAGGACGA GGAGAAGGCC GTGGACATCG
1351 TCCCTAAGTC COTATGGAAG CCGCCGCCCG TCATCCCAAG GAGGAGAGGC
1401 CGCTCAGGCG CCAACAACTA CCTGTACTTT GTGTCAACG AGCCGGCCCT
1451 GCCATGGAGC CGGCTGCCCC ACCTCACTCC AGCCAGATG GTGACGCCCC
1501 GAAAGATCAA GAAGTTCTTC ACAGGCTACC TGGACACGCC AGTCGTGAG
1551 TACCCACCCCT TCCCGGGCAA CGAGGCCAAC TACCTCGGCG CCCAGATAGC
1601 CCGCATCTCG GCGGCCACGC AGGTCAAGCC GCTGGGCTTC TACCAGTTTA
1651 GTGAGGAGGA GGGCGACGAG GAGGAGGAGG GTGCTGCTGG GCGGACTCC
1701 TACGAGGAGA ACCCGGACTT CGAGGGGATC CCCGTGCTGG AGCTGGTCTGA
1751 CTCCATGGCC AACTGGGTTG ATCACACACA GCACATCCTG CCGGAGGGCG
1801 GCTGCATCTG GGTGAACCTT TTGCAGAGA CAGAGGAGGA GGAGGACCTG
1851 GGGGAGGAGG AAGAGAAGGC AGATGAGGGG CCAGAGGAGG TGGAGCAGGA
1901 GGTGGGCCCC CCACTGCTAA CGCCACTTTC AGAAGATGGA GAAATCATGC
1951 ACCTGGCACG CTGGACCAAC CGCCTGTCTT GCAGCCTCTG CCCGCTATGC
2001 TCAGTGGCCG TTGTGCGCTC CAACCTCTGG CCGGGGGGCT ATGCTATGC
2051 CAGTGGCAAA AAGTTTGAGA ACATCTACAT CGGCTGGGGT CACAAGTAC
2101 GCGCCGAGAG CTTCACACCG GCCCTGCCAG CCGCCATTCA CACAAGTAC
2151 CCCAGTGGCC CAGAGATCAT GGAGATGAGT GACCCACAG TGGAGAGGA
2201 GCAGGCTCTG AAAGCAGCCC CCGAACAGCG CTTGGGAGCC ACAGGAGGAG
2251 AGGAGGAGGG CGAGGAGGAG GAGGAGGGCG AGGAGACAGA TGACTGAGGC
```

2301 CCACCTCTA GCCACTTTCC CCAAGCAGGT AGATAGCAAA TTTCCTCTTA  
2351 GAGGTAGTTA GCATGGATTA TATTTTCACT ATGTGCTTCC TGTCCTCCAGA  
2401 GGGCAGGGAT AGAAAAGGAA GGCAACTGCT TCAAAATAAAA TTCCTCCAGC  
2451 GCATTAAAAA AAAAAAAGG AAAAAAAG

## BLAST Results

No BLAST result

## Medline entries

86251010:

Molecular cloning and expression of flagellar radial spoke and dynein  
genes of  
Chlamydomona

81142496:

Radial spokes of Chlamydomonas flagella: polypeptide composition and  
phosphorylation of  
stalk components.

9450971:

Molecular cloning and characterization of a radial spoke head protein of sea urchin sperm  
axonemes: involvement of the protein in the regulation of sperm motility.

## Peptide information for frame 3

ORF from 144 bp to 2294 bp; peptide length: 717  
Category: strong similarity to known protein

1 MGDLPYPYPER PAQQPPGRRT SQASQRRHSR DQAQALAADP EERQIIPDA  
51 QRNAPGWSQR GSLSQQENLL MPQVFQAEAA RLGGMEYPSV NTGFFPSEFQP  
101 QPYSDESRMQ VAELTTSML QRLQQGQSSL FQQLDPTFQE PPVNPGLQFM  
151 LYQTDQFSEG AQHGPIYRDD PALQFLPSEL GFPHYSAQVP EPEPLELAVQ  
201 NAKAYLLQTS INCOLSLYEH LVNLLTKILN QRPEDPLSVL ESLNRTTQWE  
251 WFHPKLOTLR DDPEMQPTYK MAEKQKALFT RSGGGTEGEQ EMEEVGETP  
301 VPNIMETAFY FEQAGVGLSS DESFRIFLAM KQLVEQQPIH TCRFMGKILG  
351 IKRSYLVAEV EFREGEEAE EEEVEEMTEG GEVMEAHGEE EGEEDEEKAV  
401 DIVPKSVWKP PPVIPKEESR SGANKYLYFV CNEPGLPWTR LPHVTPAQIV  
451 NARKIKKFFT GYLDPVVSYS PFFPGNEANY LRAQIARISA ATQVSPGLFY  
501 QFSEEEGDEE EGGAGRDSY EENPOFEGIP VLELVDSMAN VVHHTQHILP  
551 QGRCTWVNPL QKTEEEEDLG EEEKADGEP EEVEQEVGPP LLTPLSEDAE  
601 IMHLAPWTTT LSCSLCPQYS VAVVRSNLWP GAYAYASGKK FENIYIGWGH  
651 KYSPESFNPA LPAPIQQEYP SGPEIMEMSD PTVEEEQALK AAQEALGAT  
701 EEEEGEETDD EGEETDD

## BLASTP hits

Entry U73123.1 from database TREMBL:

product: "radial spokehead"; Strongylocentrotus purpuratus radial  
spokehead mRNA, complete cds.  
Score = 1604, P = 7.4e-165, identities = 303/523, positives = 395/523

Entry B44498 from database PIR:

radial spoke protein 6 - Chlamydomonas reinhardtii  
Score = 386, P = 3.4e-45, identities = 105/264, positives = 138/264

Alert BLASTP hits for DKFZphtes3\_15i5, frame 3

No Alert BLASTP hits found

Pedant information for DKFZphtes3\_15i5, frame 3

## Report for DKFZphtes3\_15i5.3

[LENGTH] 717  
[MW] 80913.61  
[pI] 4.36

```

(HOMOL)      TREMBL:U73123_1 product: "radial spokehead"; Strongylocentrotus purpuratus
radial spokehead mRNA, complete cds. 1e-130
[PROSITE]    TRANSFERRIN_1      1
[PROSITE]    MYRISTYL           5
[PROSITE]    AMIDATION          2
[PROSITE]    CAMP_PHOSPHO_SITE  2
[PROSITE]    CK2_PHOSPHO_SITE   14
[PROSITE]    TYR_PHOSPHO_SITE   1
[PROSITE]    GLYCOSAMINOGLYCAN  1
[PROSITE]    PKC_PHOSPHO_SITE   8
[PROSITE]    ASN_GLYCOSYLATION  1
[KW]         All_Alpha
[KW]         LOW_COMPLEXITY     21.48 %

```

```

SEQ  MGDLPYPYPERPAQPPGRRTSQASQRRHSRDQAQALAADPEERQQIPDQARNAPGWSQR
SEG  .....XXXXXXXXXXXXX.....
PRD  cccccccccccccccccccccchhhhhhhhhhhhhhhhhhhhhcccccccccccccccccccc

SEQ  GSLSQENLLMPQVFQAEEARLGGMEYPSVNTGFPSEFQPPQYSDSRMQVAELTTSML
SEG  .....XXXXXXXXXXXXX.....
PRD  cccchhhhhhhhhhhhhhhhhhhhhccccccccccccccccccccccccchhhhhhhhhhhhhhh

SEQ  QRLQOQGSSLFQQLDPTFQEPVFNPLGQFNLYQTQDFSEGAQHGPYIRDDPALQFLPSEL
SEG  .....XXXXXXXXXXXXX.....
PRD  hhhhcccccccccccccccccccccccccccccccccccccccccccccccccccccccc

SEQ  GFPHYSAQVPEPEPELEAVQNAKAYLLQTSINCDLSLYEHLVNLTKILNORPEDPLSVL
SEG  .....
PRD  cccccccccccccchhhhhhhhhhhhhhhhhhhhhccccchhhhhhhhhhhhhhhhhhhccccchhh

SEQ  ESLNRTTQWEFHPKLDLTDPEMQPTYKMAEKQKALFTRSGGGTEGEQEMEEVEGETP
SEG  .....XXXXXXXXXXXXX.....
PRD  hhhchhhhhccccccccccccccccccccchhhhhhhhhhhhhhhhhhhhhccccchhhhhhhhhcccc

SEQ  VPMIMETAFYFEQAGVGLSSDESFRILAMKQLVEQQPIHTCRFWGKILGIRSYLVAEV
SEG  .....
PRD  cccchhhhhhhhhhhccccccccchhhhhhhhhhhhhhhhhhhhhccccchhhhhhhhhccccchhhhh

SEQ  EFREGEEEEEVEEMTEGGEVMEAHGEEEGEEDDEEKAVDIVPKSVWKPVPPIPKESR
SEG  .....XXXXXXXXXXXXX.....
PRD  hhhhhhhhhhhhhhhhhhhhhccccccccccccccccccccchhhheeecccccccccccccccccc

SEQ  SGANKLYFVCNEPGLPWTRLPHVTQAQIVNARKIKKFTGYLDPVVSYPFPFGNEANY
SEG  .....
PRD  cccceeeeeccccccccccccccccchhhhhhhhhhhhhhhhhhhhhccccccccccccccccchhh

SEQ  LRAQIARISAATQVSPGLGFYQFSEEGDEEEGGAGRDSYEENPDFEGIPVLELVDSMAN
SEG  .....XXXXXXXXXXXXX.....
PRD  hhhhhhhhhhhhhccccccccccccccccccccccccccccccccccccccccccccchhhhh

SEQ  WVHTQHILPQGRCTWVNPQLKTEEEEDLGEEEKADEGPVEVEQEVGPPLTPLSEDAE
SEG  .....XXXXXXXXXXXXX.....
PRD  hhhccccccccccccccccchhhhhhhhhhhccccchhhhhcccccccccccccccccccccccc

SEQ  IMHLAPWTTRLSCSLCPQYSVAVVRSNLWPGAYAYASGKKFENIYIGWGHKYSPEFNP
SEG  .....
PRD  ccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc

SEQ  LPAPIQEQYPSGPEIMEMSDPTVEEQALKAQEQALGATEEEEEEGEEEEEGETDD
SEG  .....XXXXXXXXXXXXX.....
PRD  cccccccccccccccccccccchhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh

```

#### Prosite for DKF2phtes3\_15i5.3

PS00001	244->248	ASN_GLYCOSYLATION	PDOC00001
PS00002	282->286	GLYCOSAMINOGLYCAN	PDOC00002
PS00004	18->22	CAMP_PHOSPHO_SITE	PDOC00004
PS00004	26->30	CAMP_PHOSPHO_SITE	PDOC00004
PS00005	24->27	PKC_PHOSPHO_SITE	PDOC00005
PS00005	58->61	PKC_PHOSPHO_SITE	PDOC00005
PS00005	258->261	PKC_PHOSPHO_SITE	PDOC00005
PS00005	268->271	PKC_PHOSPHO_SITE	PDOC00005
PS00005	323->326	PKC_PHOSPHO_SITE	PDOC00005
PS00005	341->344	PKC_PHOSPHO_SITE	PDOC00005
PS00005	608->611	PKC_PHOSPHO_SITE	PDOC00005
PS00005	637->640	PKC_PHOSPHO_SITE	PDOC00005
PS00006	64->68	CK2_PHOSPHO_SITE	PDOC00006
PS00006	137->141	CK2_PHOSPHO_SITE	PDOC00006

WO 01/12659

PCT/IB00/01496

PS00006	216->220	CK2_PHOSPHO_SITE	PDOC00006
PS00006	238->242	CK2_PHOSPHO_SITE	PDOC00006
PS00006	247->251	CK2_PHOSPHO_SITE	PDOC00006
PS00006	258->262	CK2_PHOSPHO_SITE	PDOC00006
PS00006	286->290	CK2_PHOSPHO_SITE	PDOC00006
PS00006	319->323	CK2_PHOSPHO_SITE	PDOC00006
PS00006	503->507	CK2_PHOSPHO_SITE	PDOC00006
PS00006	519->523	CK2_PHOSPHO_SITE	PDOC00006
PS00006	563->567	CK2_PHOSPHO_SITE	PDOC00006
PS00006	671->675	CK2_PHOSPHO_SITE	PDOC00006
PS00006	682->686	CK2_PHOSPHO_SITE	PDOC00006
PS00006	700->704	CK2_PHOSPHO_SITE	PDOC00006
PS00007	639->646	TYR_PHOSPHO_SITE	PDOC00007
PS00008	284->290	MYRISTYL	PDOC00008
PS00008	315->321	MYRISTYL	PDOC00008
PS00008	350->356	MYRISTYL	PDOC00008
PS00008	435->441	MYRISTYL	PDOC00008
PS00008	475->481	MYRISTYL	PDOC00008
PS00009	16->20	AMIDATION	PDOC00009
PS00009	637->641	AMIDATION	PDOC00009
PS00205	619->628	TRANSFERRIN_1	PDOC00182

(No Pfam data available for DKF2phtes3\_15i5.3)

DKF2phtes3\_15j18

group: testes derived

DKF2phtes3\_15j18 encodes a novel 148 amino acid protein without similarity to known proteins.

No informative BLAST results; No predictive prosite, pfam or SCOP motif.

The new protein can find application in studying the expression profile of testis-specific genes.

unknown

complete cDNA, complete cds, few EST hits

Sequenced by GBF

Locus: unknown

Insert length: 905 bp

Poly A stretch at pos. 839, polyadenylation signal at pos. 815

```
1 GTGATTTCATA TGCTTCCATA GCAGGTGTCT GCTTCTGAGC CAAGCTCCCA
51 GGGCAGCGGA GCAGGCACCA ACCAGCATCC CAGGGGAGGG CACAGCTTGT
101 CCAGCTGGGA TGTTTGGGTG CCCTGTGAGA TGCCCCAAGC CACCAACCCA
151 GCTTATCTCA GGAGAAGCCT CGCGGCCCGG TGTGCCGCCG TGGAGAGATG
201 TGCTACAGCA GCCGGGGGTG GGGGGAGAGG GTGGGCTTAG ARTCTCTTGG
251 CAGGGAGCCC CCAAGAGCAG GGTGAGACCT GCCTTCATT CACCTGTCCC
301 CTTACAGTT CTGCAAAGCC AGCATTATCA TCCCTTTTCA GAAGGAGTGG
351 GCACCTCAGT GGAATGCCCT ACCCCAGTCC TGGGCTGGA AAGCGATATG
401 GCCAGGACTG CACCCACCCC CTCATCCCTG CACCCTTCC CTGCCTGGGA
451 TTCTCCAGC CCTGTGCACT GTGGAGCGCC TGTGCTTCC GCTCATGGAG
501 GTTTCCCAAG GGCACGCGCT GAGGGCAGCT GGTCTCAGCC TGGGGCCGGG
551 TCCTAGTAAC TGCTCTCTT TGCTTTCCAG CCAGTGTITT GGGGTTTGAA
601 GTTGAATCT TCAGCTACTG TCAAGAACAG CCACAAAAAT GTGTCACGAT
651 CAAGATCTTT GAGAGTCCAC CAATCAGGAG GCGTCTGTGA CAGTCGCTGT
701 CTTCTCAGAA CAGAATCCAC ACCCAGGATT CAACCCAAAT GATTCTCAT
751 CAGGTGATTC TTGTTGTAG CAAAGTTTAT GTGAATGTGG GTGAGTTTCT
801 GTTATGAATG TGGTCAATAA ATGTTATTGG TGAACCTCTA AAAAAAAAAA
851 AAAAAAAAAA GGGCGCCGCT CTAGAGGATC CAAGCTTACG TACGCGAAAA
901 AAAAG
```

## BLAST Results

No BLAST result

## Medline entries

No Medline entry

## Peptide information for frame 2

ORF from 110 bp to 553 bp; peptide length: 148  
Category: putative protein

```
1 MFGCPVRCPK PPTQLISGEA SAARLPAPWRD VLQQPGVGGE GGLRISWQGA
51 PKSRVRPAFI SPVPFTVLQS QHYHPFSEGV GTQVECLTPV LRLESDMART
101 AHPSSSLHPF PAWDSSSPVH CGAPLPSAHG GFPRARAEGS WSQPGAGS
```

## BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKF2phtes3\_15j18, frame 2

No Alert BLASTP hits found

Pedant information for DKF2phtes3\_15j18, frame 2





DKF2phtes3\_15j3  
-----

group: nucleic acid management

DKF2phtes3\_15j3 encodes a novel 743 amino acid protein with similarity to proteins with unknown function.

The novel protein contains a RNA recognition motif, predicted by Pfam and therefore binds to RNA. The protein is similar to YGR276C, a ribonuclease H of *S. cerevisiae*. Thus, the protein seems to be a new RNA-modifying protein.

The new protein can find application in modulating the RNA metabolism in human cells and as a tool for biotechnologic manipulations.

"44M2.3"; product, differences to genmodel, similarity to ribonuclease H

complete cDNA, complete cds, EST hits  
YGR276C = ribonuclease H  
differences to genmodel of 44M2.3

Sequenced by GBF

Locus: /map="16p11.2"

Insert length: 2695 bp

Poly A stretch at pos. 2601, polyadenylation signal at pos. 2579

```
1 GCGGTTGTTG TTGGCAGCTG TGGCTAAGGA GGGGAGAACC TCTGCTCCCC
51 GCCCGTCTTC TCTTCTGCGT TTCCCGGGCT AGGGGGCGTG GGGAGTGGTT
101 TTAGGCGGCG AAGCCGCTCG GCAGCACCTT CTTCTTTGCG CAGGCAGACG
151 CCCGTTGTAG CCGTTGGGGA ACCGTTGAGA ATCCGCCATG GAGCCAGAGA
201 GGGAAAGGGAC CGAGAGACAC CCCAGGAAGG TCAGGGAAAG CAGGCAGGCC
251 CCAATAAAGC TGGTCGGGGC AGCTGAGGCG ATGAAAGCCG GTTGGGATCT
301 CGAGGAGAGT CAGCCCGAGG CCAAGAAAGC CCGCTTATCT ACCATTTTAT
351 TTATGACAAA CTGTGAAGTA ACCCATGACC AGCTGTGTGA ATTGCTGAAG
401 TATGCACTTC TGGGCAATC CAATGTTCCA AAACCCAGCT GGTGCCAGCT
451 TTTTCATCAA AACCACTAA ACAACGTAGT GGTTTTGTGT CTGCAGGGAA
501 TGAGTCAGCT ACACTTTAC AGGTTCTATT TGGAGTTTGG ATGCTTTCGA
551 AAAGCATTCG GACATAAAT CCGCTTGCTC CCACCATCAT CTGATTTTCT
601 AGCTGATGTT GTTGGGCTAC AAACCTGAAC AAGAGCTGGA GATCTGCCCA
651 AGACAATGGA AGGGCCTTTA CCTTCTAATG CAAAAGCCGC CATCAACCTT
701 CAGGATGATC CCATCATTCA AAAGTATGGC TCTAAGAAAG TGGGCTTGAC
751 CAGATGCCTT CTGACAAAGG AGGAAATGAG AACGTTTCAC TTTCCATTAC
801 AAGGTTTTCG TGATTGTGAA AACTTTTAC TTACCAATG TAATGGTTCT
851 ATAGCAGACA ATAGTCCTCT CTTTGGACTT GACTGTGAAA TGTGCTTCAC
901 ATCCAAGGGG AGAGAGCTAA CACGCATCTC ACTGGTTGCT GAAGGAGGCT
951 GCTGTGTTAT GGATGAAGT GTCAAACTCG AAAACAAGAT TCTGGACTAC
1001 CTCACCAAGT TTTGGGAAT CACGAAGAAG ATTCTTAACC CAGTGACGAC
1051 CAAACTCAAA GATGTACAGA GGCAGTTAAA AGCACTGCTT CCTCTGATG
1101 CTGTGTTAGT GGGCCACTCC TTAGATTGGG ATCTCAGAGC ACTGAAATG
1151 ATACATCCAT ATGTTATTGA TACATCGTTG CTTTATGTCA GAGAGCAGGG
1201 CAGAAGATTT AAGCTCAAGT TCTTAGCCAA AGTTATTTTG GGAAGGATA
1251 TACAGTGTCG AGACAGACTT GGTCTAGTAT CCACAGAAGA TGCTAGAACA
1301 ATCCTTGAAT TGGCTCGGTA TTTCTTAAAG CATGGCCCAA AAAAGATTGC
1351 AGAACTAAAT CTAGAAGCAC TAGCTAATCA CCAAGAAATA CAAGCAGCAG
1401 GCCAAGAGCC TAAAAACACA GCAGAAGTAC TTCAGCACCC AAACACAAGT
1451 GTTTTAGAAT GCTTGGATTC AGTGGGTCAG AAGCTTCTTT TTTTGACCCG
1501 GGAGACAGAT GCTGGTGAAC TTCCATCTTC CAGAAATTGT CAAACTATTA
1551 AGTGCTTTTC AAATAAAGAG GTTCTTGAGC AGGCCAGAGT GGAATCCCC
1601 CTGTTTCCCT TCAGCATTGT TCAGTTCTCT TTTAAGCCCT TTTACCTGT
1651 CCTCACTGAG GAGATGAACA AAAGGATGAG GATCAAGTGG ACAGAGATAT
1701 CAACTGTCTA TGCTGGGCCA TTTAGCAAAA ATTGCAATCT CAGGGCTCTG
1751 AAGAGGCTGT TTAAGAGCTT TGGCCGAGTC CAGTCAATGA CTTTGTGTTCT
1801 TGAACCCCGT CAGGTGCAGA GGCCTGTGAC AGAGCTCAGC CTTGATTGTG
1851 ACACCCCTCG GAATGAGCTG GAAGGAGATT CTGAAAAACA AGGCTCTATA
1901 TATCTGTCTG GAGTGAGTGA AACCTTCAAA GAACAGCTAT TGCAGGAGCC
1951 CCGCCTCTTT CTGGCCCTGG AAGCTGTGAT CTTGCCATAA GATCTTAAAA
2001 GTGGAAAGCA GAAAAAATAC TGTTTCTGTA AATTCAAAG TTTGGCAGT
2051 GCCCAGCAGG CCTCAACAT TCTCACAGGC AAGGACTGGA AGCTGAAAGG
2101 CAGGCATGCC CTAACCCCA GGCACCTCCA TGCCCTGGCTC AGAGGCTTAC
2151 CACCTGAATC AACAGGCTC CCAGGGCTTC GTGTTGTACC TCCCCCTTT
2201 GAACAGGAGG CTTGCAGAC TCTGAAACTG GACCACCCGA AGATAGCAGC
2251 CTGGCGCTGG AGCCGGAAGA TTGGAAGACT CTACACAGC TTGTGCCCGG
2301 GCACTCTCTG CCTCATCTG CTGCCAGGAA CCAAGAGCAC TCATGGTTCA
2351 CTCTCTGGTC TAGGACTGAT GGGAAATAAA GAGGAAGAAG AAAGCGCTGG
2401 CCCAGGCTG TGTTCGTGAG TCGGCCTGCC ATGTTTCCAT GTGCCATTTC
```

```
2451 TTACCCCTTG TAGGCAATGG CAAAGAATGT GGTACGGCTG TAGCCTCCCC
2501 AACCAGCAGA CAGTTTTATG GAAACTTGGT ATAGCAGCTA AAAGAGTTTA
2551 GTTTGTTTAT ATGGCATGTA TAAGTTTCA ATAAATGCCT AAAGTTCAAG
2601 CATAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA
2651 AGGCGCGCCG CTCTAAAGGA TCCAAGCTTA CGTACGCGAA AAAAG
```

## BLAST Results

No BLAST result

## Medline entries

No Medline entry

## Peptide information for frame 2

ORF from 188 bp to 2416 bp; peptide length: 743  
Category: similarity to known protein

```
1 MEPEEREGTER HPRKVRRESQ APNKLVGAAE AMKAGWDL EE SQPEAKKARL
51 STILFTDNCE VTHDQLCELL KYAVLGKSNV PKPSWCQLFH QNHLNNVVVF
101 VLQGMSQLHF YRFYLEFGCL RKAFRHKFRL PPPSSDFLAD VVGLQTEQRA
151 GDLPKTMEGP LPSNAKAAIN LQDDPIIQKY GSKKVLTRC LLTKEEMRTF
201 HFPLQGFPDC ENFLLTKCNG SIADNSPLFG LOCEMCLTSK GRELTRISLV
251 AEGGCCVME LVKPENKILD YLTSFSGITK KILNPVTTKL KDVRQLKAL
301 LPPDAVLVGH SLDLDLRALK MIHPYVIDTS LLYVREQGRR FKLKFLAKVI
351 LGKDIQCPDR LGHDATEDAR TLELARYFL KHGPKKIAEL NLEALANHQE
401 IQAAGQEPKN TAEVLQHPNT SVLECLDSVG QKLLFLTRET DAGELPSSRN
451 CQTIKCLSNK EVLEQARVEI PLFPFSIVQF SFKAFSPVLT EEMNKRMRK
501 WTEISTVYAG PFSKNCNLRA LKRLFKSFGP VQSMTFVLET RQVQRPVTEL
551 TLQCDTLVNE LEGDSENQGS IYLSGVSETF KEQLLQEPRL FLGLEAVILP
601 KDLKSGKQKK YCFLKFKSFG SAQQALNILT GKDWKLGRRH ALTPRHLHAW
651 LRGLPPESTR LPGLRVVPPP FEQALQTLK LDHPKIAAMR WSRKIGKLYN
701 SLCPTGLCLI LLPGTKSTHG SLSCGLGLMI KEEESAGPG LCS
```

## BLASTP hits

No BLASTP hits available

## Alert BLASTP hits for DKFZphtes3\_15j3, frame 2

TREMBL:AC004381\_4 gene: "44M2.3"; product: "Unknown gene product";  
Homo sapiens Chromosome 16 BAC clone CIT987SK-44M2, complete sequence.,  
N = 2, Score = 1827, P = 2.1e-284

TREMBL:AF016430\_4 gene: "C05C8.5"; Caenorhabditis elegans cosmid  
C05C8., N = 2, Score = 370, P = 1.7e-34

PIR:S64609 hypothetical protein YGR276c - yeast (Saccharomyces  
cerevisiae), N = 2, Score = 334, P = 1.8e-27

TREMBLNEW:SPAC637\_9 gene: "SPAC637.09"; product: "putative  
exonuclease"; S.pombe chromosome I cosmid c637., N = 3, Score = 326, P  
= 2.8e-27

>TREMBL:AC004381\_4 gene: "44M2.3"; product: "Unknown gene product"; Homo  
sapiens Chromosome 16 BAC clone CIT987SK-44M2, complete sequence.  
Length = 547

## HSPs:

Score = 1827 (274.1 bits), Expect = 2.1e-284, Sum P(2) = 2.1e-284  
Identities = 358/373 (95%), Positives = 358/373 (95%)

Query: 105 MSQLHFYRFYLEFGCLRKAFRHKFRLPPPSSDFLADVVLQTEQRAGDLPKTMEGPLPSN 164  
MSQLHFYRFYLEFGCLRKAFRHKFRLPPPSSDFLADVVLQTEQRAGDLPKTMEGPLPSN  
Sbjct: 1 MSQLHFYRFYLEFGCLRKAFRHKFRLPPPSSDFLADVVLQTEQRAGDLPKTMEGPLPSN 60

Query: 165 AKAAINLQDDPIIQKYGSKKVLTRCLLTKEEMRTFHFLQGFDPDCENFLLTKCNGSIAD 224  
AKAAINLQDDPIIQKYGSKKVLTRCLLTKEEMRTFHFLQGFDPDCENFLLTKCNGSIAD  
Sbjct: 61 AKAAINLQDDPIIQKYGSKKVLTRCLLTKEEMRTFHFLQGFDPDCENFLLTKCNGSIAD 120

```

Query: 225 NSPLFGLDCM-----CLTSKGRELTRISLVAEGGCCVMDLVKPKENKIL 269
      NSPLFGLDCM          CLTSKGRELTRISLVAEGGCCVMDLVKPKENKIL
Sbjct: 121 NSPLFGLDCMARTTFNFSIGVLQAECLTSKGRELTRISLVAEGGCCVMDLVKPKENKIL 180

Query: 270 DYLTFSFGITKKILNPVTTKLKDVQRQKALLPPDAVLVGHSLDLDLRALKMHPYVIDT 329
      DYLTFSFGITKKILNPVTTKLKDVQRQKALLPPDAVLVGHSLDLDLRALKMHPYVIDT
Sbjct: 181 DYLTFSFGITKKILNPVTTKLKDVQRQKALLPPDAVLVGHSLDLDLRALKMHPYVIDT 240

Query: 330 SLLYVREQGRRFKLFLAKVILGKDIQCPDRLGHDAEDARTILELARYFLKHGPKKIAE 389
      SLLYVREQGRRFKLFLAKVILGKDIQCPDRLGHDAEDARTILELARYFLKHGPKKIAE
Sbjct: 241 SLLYVREQGRRFKLFLAKVILGKDIQCPDRLGHDAEDARTILELARYFLKHGPKKIAE 300

Query: 390 LNLEALANHQEIQAAQGEKNTAEVLQHPNTSVLECLDSVGQKLLFLTRETDA GELPSSR 449
      LNLEALANHQEIQAAQGEKNTAEVLQHPNTSVLECLDSVGQKLLFLTRETDA GELPSSR
Sbjct: 301 LNLEALANHQEIQAAQGEKNTAEVLQHPNTSVLECLDSVGQKLLFLTRETDA GELPSSR 360

Query: 450 NCQTIKCLSNKEV 462
      NCQTIKCLSNKEV
Sbjct: 361 NCQTIKCLSNKEV 373

Score = 929 (139.4 bits), Expect = 2.1e-284, Sum P(2) = 2.1e-284
Identities = 175/179 (97%), Positives = 177/179 (98%)

Query: 538 LETRQVQRPVTELTLDLDCDTLVNELEGDSNQGSIYLSGVSETFKEQLLQEPRLFGLGLEAV 597
      L ++VQRPVTELTLDLDCDTLVNELEGDSNQGSIYLSGVSETFKEQLLQEPRLFGLGLEAV
Sbjct: 368 LSNKEVQRPVTELTLDLDCDTLVNELEGDSNQGSIYLSGVSETFKEQLLQEPRLFGLGLEAV 427

Query: 598 ILPKDLKSGKQKYYCFLKFKSFGSAQQALNLTGKDWKLGKRGHALTPRHLHAWLRGLPPE 657
      ILPKDLKSGKQKYYCFLKFKSFGSAQQALNLTGKDWKLGKRGHALTPRHLHAWLRGLPPE
Sbjct: 428 ILPKDLKSGKQKYYCFLKFKSFGSAQQALNLTGKDWKLGKRGHALTPRHLHAWLRGLPPE 487

Query: 658 STRLPGLRVVPPPFEEQALQTLKLDHPKIAAWRWSRKIGLYNSLCPGTLCLILLPGTK 716
      STRLPGLRVVPPPFEEQALQTLKLDHPKIAAWRWSRKIGLYNSLCPGTLCLILLPGTK
Sbjct: 488 STRLPGLRVVPPPFEEQALQTLKLDHPKIAAWRWSRKIGLYNSLCPGTLCLILLPGTK 546

```

Pedant information for DKFZphtes3\_15j3, frame 2

#### Report for DKFZphtes3\_15j3.2

```

[LENGTH]      743
[MW]           83536.58
[pI]           8.87
[HOMOL]        TREMBL:AC004381.4 gene: "44M2.3"; product: "Unknown gene product"; Homo sapiens
Chromosome 16  BAC clone CIT9875K-44M2, complete sequence. 0.0
[FUNCAT]       01.03.16 polynucleotide degradation [S. cerevisiae, YGR276c] 4e-30
[PROSITE]      99 unclassified proteins [S. cerevisiae, YLR107w] 3e-13
[PROSITE]      05.04 translation (initiation, elongation and termination) [S. cerevisiae,
YGL094c] 1e-10
[PROSITE]      04.05.05 mrna processing (5'-end, 3'-end processing and mrna degradation) [S.
cerevisiae, YGL094c] 1e-10
[PROSITE]      03.22 cell cycle control and mitosis [S. cerevisiae, YOL080c] 2e-10
[PROSITE]      MYRISTYL 5
[PROSITE]      AMIDATION 1
[PROSITE]      CK2_PHOSPHO_SITE 8
[PROSITE]      TYR_PHOSPHO_SITE 1
[PROSITE]      GLYCOSAMINOGLYCAN 1
[PROSITE]      PKC_PHOSPHO_SITE 16
[PROSITE]      ASN_GLYCOSYLATION 2
[PROSITE]      RNA recognition motif. (aka RRM, RBD, or RNP domain)
[PROSITE]      Alpha_Beta

```

```

SEQ  MEPEREGTERHPRKVRRESRQAPNKLVGAAEAMKAGWOLEESQPEAKKARLSTILFTDNCE
PRD  cchhhhhccccchhhhhhhhhcchhhhhhhhhccccccccchhhhhcccccccccc

SEQ  VTHQQLCELLKYAVLGSNVPKPSWCQLFHQNHNNVNVVFLQGMSQLHFYRFLYFGL
PRD  eehhhhhhhhhhhhhhhhhccccccccceeeccccccccceeeccccccccchhhhhhhhhhhhhhhhh

SEQ  RKAFRHKFRLPPSSDFLADVGLQTEQRAGDLPKTMEGPLPSNARAAINLQDDPIIQKY
PRD  hhhhhhhhhccccccccchhhhhhhhhhhhhhhhhccccccccccccchhhhhhhhhcccccccc

SEQ  GSKKVGTLRCLLTKEEMRTFHFPLQGFDCENFLLTKNGSIADNSPLFGLDCMCLTSK
PRD  cccccchhhhhhhhhhhhhhhhhcccccccccccccccccccccccccccccccccccccccccccc

SEQ  GRELTRISLVAEGGCCVMDLVKPKENKILDYLTFSFGITKKILNPVTTKLKDVQRQKAL
PRD  cchhhhhheeeccccccccccccccccccccccccccccccccccccccccchhhhhhhhhhhhhhhhh

```

[illegible]

Prosites for DKFZphtes3\_15j3.2

PS000001	219->223	ASN_GLYCOSYLATION	PDCC000001
PS000001	419->423	ASN_GLYCOSYLATION	PDCC000001
PS000002	723->727	GLYCOSAMINOGLYCAN	PDCC000002
PS000005	8->11	PKC_PHOSPHO_SITE	PDCC000005
PS000005	182->185	PKC_PHOSPHO_SITE	PDCC000005
PS000005	238->241	PKC_PHOSPHO_SITE	PDCC000005
PS000005	279->282	PKC_PHOSPHO_SITE	PDCC000005
PS000005	287->290	PKC_PHOSPHO_SITE	PDCC000005
PS000005	447->450	PKC_PHOSPHO_SITE	PDCC000005
PS000005	453->456	PKC_PHOSPHO_SITE	PDCC000005
PS000005	458->461	PKC_PHOSPHO_SITE	PDCC000005
PS000005	481->484	PKC_PHOSPHO_SITE	PDCC000005
PS000005	579->582	PKC_PHOSPHO_SITE	PDCC000005
PS000005	605->608	PKC_PHOSPHO_SITE	PDCC000005
PS000005	630->633	PKC_PHOSPHO_SITE	PDCC000005
PS000005	643->646	PKC_PHOSPHO_SITE	PDCC000005
PS000005	658->661	PKC_PHOSPHO_SITE	PDCC000005
PS000005	678->681	PKC_PHOSPHO_SITE	PDCC000005
PS000005	692->695	PKC_PHOSPHO_SITE	PDCC000005
PS000006	41->45	CK2_PHOSPHO_SITE	PDCC000006
PS000006	193->197	CK2_PHOSPHO_SITE	PDCC000006
PS000006	221->225	CK2_PHOSPHO_SITE	PDCC000006
PS000006	371->375	CK2_PHOSPHO_SITE	PDCC000006
PS000006	421->425	CK2_PHOSPHO_SITE	PDCC000006
PS000006	458->462	CK2_PHOSPHO_SITE	PDCC000006
PS000006	579->583	CK2_PHOSPHO_SITE	PDCC000006
PS000006	630->634	CK2_PHOSPHO_SITE	PDCC000006
PS000007	370->379	TYR_PHOSPHO_SITE	PDCC000007
PS000008	27->33	MYRISTYL	PDCC000008
PS000008	186->192	MYRISTYL	PDCC000008
PS000008	575->581	MYRISTYL	PDCC000008
PS000008	714->720	MYRISTYL	PDCC000008
PS000008	720->726	MYRISTYL	PDCC000008
PS000009	337->341	AMIDATION	PDCC000009

Pfam for DKFZphtes3\_15j3.2

HMM_NAME	RNA recognition motif. (aka RRM, RBD, or RNP domain)	
HMM	*IVXGNLPWDTTEEDLRDLfSQQFpIvsIRMMDRrETRSGRGAfVEFED	
	IV+ ++ + +E+L + + F + + + + + + + + G + + + F + +	
Query	571 IYLSGVS-ETFKQLQLEPRFLFLGLEAVILPKDLKSGKQKCYFLKFKS	618
HMM	EEDAEaKdIRMG...meFmGrIRV* +A+ A+ + G + + GR +	
Query	619 FGSAAQALNLTGDKWKLKGRHALT	643

DKFZphtes3\_15k11

group: signal transduction

DKFZphtes3\_15k11 encodes a novel 958 amino acid protein C-terminal identical with human KIAA0781 protein and high similarity to protein kinases.

The novel protein contains a protein kinase ATP-binding region signature and a serine/threonine protein kinase active-site signature. The related murine kinase was cloned from the myocardium of the developing heart.

The new protein can find application in modulation of intracellular signal pathways dependent on this kinase.

KIAA0781, 5' extension

complete cDNA, complete cds, potential start at Bp 97, EST hits

Sequenced by GBF

Locus: /map="11"

Insert length: 4868 bp

Poly A stretch at pos. 4798, polyadenylation signal at pos. 4776

```
1 GAGCAAGCGG AGCGGCCGTC GCCCAAGCCA AGCGCGCGTG CCAACCTCC
51 CGCCCGCCCG CGCTCCTGTC CGCCGTGTCT AGCAGCGGGG CCCAGCATGG
101 TCATGGCGGA TGGCCCGAGG CACTTGCAGC GCGGGCCGCT CCGGTCGGGG
151 TTCTACGACA TCGAGGGGAC GCTGGGCAAG GGCAACTTCG CTCTGGTAA
201 GCTGGGGCGG CACCGGATCA CCAAGACGGA GGTGGCAATA AAAATAATCG
251 ATAAGTCTCA GCTGGATGCA GTGAACCTTG AGAAAACTTA CCGAGAACTA
301 CAAATAATGA AAATGTTAGA CCACCTCAC ATAATCAAA TTTATCAGGT
351 AATGGAGACC AAAAGTATGT TGTACCTTGT GACAGAATAT GCCAAAAATG
401 GAGAAATTTT TGACTATCTT GCTAATCATG GCCGGTTAAA TGAGTCTGAA
451 GCCAGGCGAA AATTCTGGCA AATCCTGTCT GCTGTTGATT ATTGTCTGG
501 TCGGAAGATT GTGCACCGTG ACCTCAAAGC TGAAAACTCT CTCTGGGATA
551 ACAACATGAA TATCAAAATA GCAGATTTCG GTTTTGAAAA TTCTTTTAAA
601 AGTGGTGAAC TGCTGGCAAC ATGGTGTGGC AGCCCCCTTT TGTGAGCCCC
651 AGAAGTCTTT GAAGGGCAGC AGTATGAAG ACCACAGCTG GACATCTGGA
701 GTATGGGAGT TGTCTTTTAT GTCTTGTCT GTGAGCTCTT GAAGATTCCG
751 GGACCGACTC TTCCAATTTT GAGGCGAGGG GCACCTTATC CGAAGGATCT
801 GATTCCGTAT TTCAATGTCAG AAGATTGCGA TAGCCCAAT CAAGCAGCAT
851 TGGTCTGTAGA CCCATCCAAA CGGCTAACCA AGACCTGTTC TCTATCCACA
901 AAATGGATGC TCATAGAAGT TCCTGTCCAG GTTATATGAG CAGGTTCTGC
951 AGAGCAGAAA AATGAGCCAT CAGCTTGGGA ATATGCTGCG AGAAAACCAT TGAGTCTTTG
1001 GACTGATGCA CAGCCTTGGG ATAGATCAGC AGTATGCTGC ATTTATTCTT TGTGGTGGGA
1051 CAGAACAAGA GCTATACGGA CTCTGCTGCC ATTTATTCTT TGTGGTGGGA
1101 GCGCCTGAAA TCACATCGGA GCAGTTTCCC AGTGGAGCAG AGACTTGATG
1151 GCGCCAGGCG TCGGCCCTAGC ACCATTCTCT AGCAACACTT TGCCAAAGGCA
1201 CAGACTGTGG GGCTCCCACT GACCATTCAT TCACCGAACA TGAGGCTGCT
1251 GCGATCTGCC CTCTCCCCCT AGCATCCAA CCGTGGAGCC TTTTCATTTC
1301 CAGCATCTGG CTGTCAAGCG GAAGCTGCAT TCATGGAAGA AGAGTGTGTG
1351 GACACTCCAA AGGTCAATGG CTGTCTGCTT GACCTGTGCT CTCCTGTCTT
1401 GGTGCGGAAG GGATGCCAGT CACTGTCCAG CAACATCATG GAGACCTCCA
1451 TTGACGAAGG GCTGGAGACA GAAGGAGAGG CCGAGGAAGA CCCCCTCAT
1501 GCCTTTGAGG CATTTAGTC CACACCGAGC GGGCAGAGAC GGCACACTCT
1551 GTCAGAAGTG ACCAATCAAC TGGTCTGTAT GCCTGGGCA GGGAAATTT
1601 TCTCCATGAA TGACAGCCCC TCCCTTGACA GTGTGGACTC TGACTATGAT
1651 ATGGGGTCTG TTCCAGAGGGG CCTGAACITT CTGGAAGACA ACCCTTCCCT
1701 TAAGGACATC ATGTTAGCCA ATCAGCCTTC ACCCCGATG ACCTTCCCT
1751 TCATAAGCCT GAGACCTACC AACCCAGCCA TGCAAGGCTC GAGCTCCAG
1801 AAACGAGAGG TCCACAACAG GTCTCCAGTG AGCTTCAGAG AGGCGCCGAG
1851 AGCATCAGAT ACCTCCCTCA CCCAGGGAAT TGTAGCATTT AGACAACATC
1901 TTCAGAAATC GGCTAGAACC AAAGGAATTC TACCTTAACC TGGGCGCGGC
1951 TTGTTGTATG AACAAATAGG ACCGGAGGCA GACCTTAACC TGGGCGCGGC
2001 GGCTCCTCAG CTCAGGAGCC TTGCTAGCAG CTGCTCTCAG GAAGAGCTTT
2051 CTCAGCAGCA GGAAGGCGTC TCCACTCTCC CTGCGAGCTT GCATCCCCCA
2101 CTGTCCCCAC GGCAGAGCCT GGAGACCCAG TACCTGCAGC ACAGACTCCA
2151 GAAGCCCGAG CTTCTGTCAA AGGCCAGAA CACCTGTGAG CTTTATTGCA
2201 AAGAACCACC GCGGAGCCTT GAGCAGCAGC TGCAGGAACA TAGGCTCCAG
2251 CAGAAGCGAC TCTTTCTTCA GAAGCAGTCT CAACCTCAGG CCTATTTTAA
2301 TCAGATGCAG ATAGCAGAGA GCTCCTACCC ACAGCCAAGT CAGCAGCTGC
2351 CCCTTCCCCG CCAGGAGACT CCACCGCCTT CTCAGCAGGC CCCACCGTTC
2401 AGCCTGACCC AGCCCTTGAG CCCCGTCTGT GAGCCTTCTC CCGAGCAGAT
2451 GCAATACAGC CTTTCTCTCA GCCAGTACCA AGAGATGCAG CTTCAGCCCC
2501 TGCCCTCCAC TTCCGGTCCC CGGGCTGTCT CTCTCTGCC CACGACGCTA
2551 CAGCAGCAGC AGCCGCCACC GCCACCAACC CTCTCACAC CACGACAGCC
2601 AGGAGCTGCC CCAGCCCTCT TACAGTTCTC CTATCAGACT TGTGAGCTGC
```

```
2651 CAAGCGCTGC TTCCCTGCG CCAGACTATC CCACTCCCTG TCAGTATCCT
2701 GTGGATGGAG CCCAGCAGAG CGACCTAACG GGGCCAGACT GTCCCAAGAG
2751 CCCAGGACTG CAAGAGGCCC CCTCCAGCTA CGACCCACTA GCCCTCTCTG
2801 AGCTACCTGG ACTCTTTGAT TGTGAAATGC TAGACCGTGT GGATCCACAA
2851 CACAACGGGT ATGTCCTGGT GAATTAGTCT CAGCACAGGA ATTGAGGTGG
2901 GTCAGGTGAA GGAAGAGTGT ATGTTCCCTA TTTTATTCCA GCCTTTTAAA
2951 TTTAAAGCTT ATTTTCTTGC CCTCTCCCTA ACGGGGAGAA ATCGAGCCAC
3001 CCAACTGGAA TCAGAGGGTC TGGCTGGGCT GGATGTTGCT TCCTCTGGT
3051 TCTGCCCCAC CACAAAGTTT TCTGTGGCAA GTGCTGGAAC ATAGTTGTAG
3101 GCTGAGGCTC CTGCCCTTCG GTCGAGTGGA GCAAGCTCTC GAGGGCAGCA
3151 CTGACAAATG TGTTCCTAAG AAGACATTCA GACCCAGGTC TTATGCAGGA
3201 TTACATCCGT TTATTATCAA GGGCAACCTT GGTGAAAGCA GAAAGGGTGT
3251 GTGCTATTGC ATATATATGG GGGAAAAGGC AATATATTTT TCAGTGAAGC
3301 TGAGCAACCA CATATTGCTA CAAGGCAAAAT CAAGAAGACA TCAGGAAATC
3351 AGATGCACAG GAAATAAAGG AAAGCTGTGC TTTGTCTTGG AATCCTAAGT
3401 TCTTAGCTGC TGATGCAAGT TGTCCCCCAA GGCCATCACA AAGCAGTGGG
3451 GCATGAGCTG TGTTCAGGG GGCCTAAAT AACAGCTGGT ACTGACCCCA
3501 GAAACCGCCT TCATCTCCAT TCGGAAGCAG GTGACACACC CCTTCAGAG
3551 GTGCCCTGGG TTGCCGAGTG TCAGAAATATA CTCAGGACTC CAGAGGTGTC
3601 ACACGTGGAA CTGACAGGAG ACCCGCCACC GTGGAGGCAG GGGGCAAGAA
3651 ACTCAAGAAC GCATCAAGAG CACCAAGCCT GGGCCAGGGA AGACAGGCTC
3701 TTCTGCACTG TTCTCGTGGG CACTGCTGGC TTGCGGGCAG TCGGTTCCA
3751 GGGTACCTGT TGTCTCTTTT CGATGTAAT AACTACTTTG ACCTTACACT
3801 ATATGTTGCT AGTAGTTTAT TGAGCTTTGT ATATTGGAG AGTTTCATAT
3851 AGGGCTTAGA GATTTTAAAG ACATGATAAA TGAACTTTTC TGTCCCATGT
3901 GAAGTGGTAG TCGGTCGCTT TTCCCCCAGA TCATGCTTTA ATTCTTTCTT
3951 TTCTGTAGAA ACCAACAGTT TCCATTATG TCAATGCTAA ATCCAAAGTC
4001 ACTTCAGAGT TTGTTTTCCA CCATGTGGGA ATCAGCATTC TTAATTTCTG
4051 TAAAGTTTGG ACTTGTAAAT AAATGTTCAA GTATTACAGC AATATTCAA
4101 GAAAGAACCA CAGATGTGTT AACCATTTAA GCAGATCATC TGCCAAACAT
4151 TATATTACTA ATAAACTTA ACCAACACTT ACAATTGAGT CATCAAAGTA
4201 AGTAAAAATT AGATGCTACA GCTAGCTAAC TGTATCCCTA GAAATGATGA
4251 ATAATTTGCC ATTTGGACAG TTAACATCCA GGTGTTACAA AGTCAGTGT
4301 AATTCCTAAG ATGATCATTT CTGCCCTTTA GAATGGCTTG TCCCATCAGC
4351 AGATGAATGT GTTAAGCACA AAGCATCTTC CTTAAAGCAC AAAGAGAGGG
4401 ACTAACTGAT GCTGCATCTA GAAACACCTT TTAAGTTGCC TTTCTCTTT
4451 GTAGTTAGCG TTCAGGCAGG TGACGTGTGG AAAGTCTAGG GGTTCCTATT
4501 CTGGCCATGC GAGCCAGCTC CCTACCAACG TCGGTAACTT GAGCAGTCCC
4551 TGTGCTGGC CAGAGACTGC CTGGTCGCCA GCGCTCACCA TGGGTGCCAG
4601 GATGCTTCGC AGAGGCACCT TGCTCACCGT TGGACTTGGT GTCAGTGGGA
4651 AAGGGCAGTG TGGGGACTGT CATTTTGTG ATTTAATAAC ACACAGTGAA
4701 AATCCAGGAA GAATGAATTA AGCTTCTTCT GGGAGTTGTT TATTCCTGCT
4751 CGTGCTTAAG ATTGATGATT TCGTGAATA AAGAATATCA TTTCAATTTA
4801 AAAAAAAGGG AAAAAAAGGG CGGCCGCTCT AGAGGATCCA AGCTTACGTA
4851 CGCGTGAATA AAAAAAAG
```

## BLAST Results

Entry HSG4921 from database EMBL:  
human STS SHGC-37164.  
Score = 1605, P = 1.9e-66, identities = 349/369

Entry AB018324 from database EMBL:  
Homo sapiens mRNA for KIAA0781 protein, partial cds.  
Score = 10725, P = 0.0e+00, identities = 2145/2145

## Medline entries

No Medline entry

## Peptide information for frame 1

ORF from the beginning to 2874 bp; peptide length: 959  
Category: known protein

```
1 EQAERPSPKP SRAANPPARP RSCPPCLAAG PSMVMADGPR HLQRPVVRVG
51 FYDIEGTLGK GNFAVVKLGR HRITKTEVAI KIIDKSQDLA VNLEKIYREV
101 QIMKMLDHPH IIKLYQVMEI KSMILYLVEY AKNGEIPDYL ANHGRNLNESE
151 ARRKFWQILS AVDYCHGRKI VHRDLKAENL LLDNNMNIKI ADFGFGNFFK
201 SGELLATWCG SPPYAAPEVF EGQQYEGPOL DIWSMGVVLY VLVCGALPFD
251 GPTLPILRQR VLEGRFRIPY FMSEDCHELI RRMLVLDPK RLTIQIKHEH
```

```

301 KMWLIEVPVQ RPVLVPQEQE NEPSIGEFNE QVLRMHSLG IDQOKTIESL
351 QNKSYNHFAA IYFLVERLK SHRSSFPVEQ RLDGRQRRPS TIAEQTVAKA
401 QTVGLPVTMH SPNMRLLSA LLPQASNVEA FSFPASGCQA EAFMEEECV
451 DTPKVNGLL DPVPPVLVRK GCQSLPSNM ETSIDEGLET EGEAEDPAH
501 AFEAFQSTRS GRRRHTLSEV TNQLVVMPGA GKIFSMNDSP SLDSVDSEYD
551 MGSVQRLNLF LEDNPSLKDI MLANQPSPRM TSPFISLRPT NPAMQALSSQ
601 KREVHNRSPV SFREGRRASD TSLTQGI VAF RQHLQNLART KGILELNKVPQ
651 LLYEQIGPEA DPNLAPAAPQ LQDLASSCPQ EEVSQQQESV STLPASVHPQ
701 LSPRQSLETQ YLQHLQKPS LLSKAQNTCQ LYCKEPPRSL EQQLQEHRLQ
751 KQRLFLQKQS QLOAYFNQMQ IAESSYPQPS QQLPLPRQET PPPSQAPPFF
801 SLTQPLSPVL EPSSEMQQYS PFLSQYQEMQ LQPLPSTSGP RAAPPLPTQL
851 QQQQPPPPPP PPPPRQPGAA PAPLQFSYQT CELPSAASPA PDYPTPCQYP
901 VDGAQQSDLT GPDCPRSPGL QEAPSSYDPL ALSELPLGLFD CEMLDVADPQ
951 HNGYVLVN

```

## BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphtes3\_15k11, frame 1

No Alert BLASTP hits found

Pedant information for DKFZphtes3\_15k11, frame 1

## Report for DKFZphtes3\_15k11.1

```

[LENGTH] 926
[MW] 103915.77
[pI] 5.70
[HOMOL] TREMBL:AB018324_1 gene: "KIAA0781"; product: "KIAA0781 protein"; Homo sapiens
mRNA for KIAA0781 protein, partial cds. 0.0
[FUNCAT] 01.05.04 regulation of carbohydrate utilization [S. cerevisiae, YDR477w]
8e-76
[FUNCAT] 11.01 stress response [S. cerevisiae, YDR477w] 8e-76
[FUNCAT] 30.03 organization of cytoplasm [S. cerevisiae, YDR477w] 8e-76
[FUNCAT] 98 classification not yet clear-cut [S. cerevisiae, YCL024w] 4e-58
[FUNCAT] 03.25 cytokinesis [S. cerevisiae, YDR507c] 3e-56
[FUNCAT] 03.04 budding, cell polarity and filament formation [S. cerevisiae, YDR507c]
3e-56
[FUNCAT] 30.02 organization of plasma membrane [S. cerevisiae, YDR122w] 1e-53
[FUNCAT] 03.22 cell cycle control and mitosis [S. cerevisiae, YKL101w] 3e-53
[FUNCAT] 30.10 nuclear organization [S. cerevisiae, YKL101w] 3e-53
[FUNCAT] 99 unclassified proteins [S. cerevisiae, YPL141c] 5e-51
[FUNCAT] 03.19 recombination and dna repair [S. cerevisiae, YPL153c] 3e-42
[FUNCAT] 03.22.01 cell cycle check point proteins [S. cerevisiae, YPL153c] 3e-42
[FUNCAT] 10.99 other signal-transduction activities [S. cerevisiae, YPL153c] 3e-42
[FUNCAT] 11.04 dna repair (direct repair, base excision repair and nucleotide excision
repair) [S. cerevisiae, YPL153c] 3e-42
[FUNCAT] 03.01 cell growth [S. cerevisiae, YFR014c] 5e-42
[FUNCAT] 03.16 dna synthesis and replication [S. cerevisiae, YMR001c] 2e-34
[FUNCAT] 03.10 sporulation and germination [S. cerevisiae, YGL180w] 1e-27
[FUNCAT] 08.13 vacuolar transport [S. cerevisiae, YGL180w] 1e-27
[FUNCAT] 06.13.04 lysosomal and vacuolar degradation [S. cerevisiae, YGL180w] 1e-27
[FUNCAT] 10.02.11 key kinases [S. cerevisiae, YBL105c] 3e-26
[FUNCAT] 04.99 other transcription activities [S. cerevisiae, YER129w] 3e-26
[FUNCAT] 02.19 metabolism of energy reserves (glycogen, trehalose) [S. cerevisiae,
YPL031c] 1e-23
[FUNCAT] 01.04.04 regulation of phosphate utilization [S. cerevisiae, YPL031c]
1e-23
[FUNCAT] 04.05.01.04 transcriptional control [S. cerevisiae, YPL031c] 1e-23
[FUNCAT] 03.13 meiosis [S. cerevisiae, YOR351c] 2e-23
[FUNCAT] 10.05.11 key kinases [S. cerevisiae, YHL007c] 8e-21
[FUNCAT] 03.07 pheromone response, mating-type determination, sex-specific proteins
[S. cerevisiae, YHL007c] 8e-21
[FUNCAT] 09.01 biogenesis of cell wall [S. cerevisiae, YPL140c] 2e-20
[FUNCAT] 10.03.11 key kinases [S. cerevisiae, YLR113w] 7e-20
[FUNCAT] 04.05.01.01 general transcription activities [S. cerevisiae, YDL108w]
3e-19
[FUNCAT] 10.05.09 regulation of g-protein activity [S. cerevisiae, YBL016w] 2e-18
[FUNCAT] 10.04.11 key kinases [S. cerevisiae, YLR362w] 3e-18
[FUNCAT] 04.03.99 other trna-transcription activities [S. cerevisiae, YOR061w]
4e-18
[FUNCAT] 06.07 protein modification (glycosylation, acylation, myristylation,
palmitoylation, farnesylation and processing) [S. cerevisiae, YFL033c] 4e-17
[FUNCAT] 05.07 translational control [S. cerevisiae, YDR283c] 2e-16
[FUNCAT] 01.02.04 regulation of nitrogen and sulphur utilization [S. cerevisiae,
YNL183c] 2e-14

```

[FUNCAT] 08.99 other intracellular-transport activities [S. cerevisiae, YNL183c]  
2e-14  
[FUNCAT] 09.04 biogenesis of cytoskeleton [S. cerevisiae, YNL020c] 5e-14  
[FUNCAT] c energy conversion [M. genitalium, MG109] 2e-12  
[FUNCAT] 30.09 organization of intracellular transport vesicles [S. cerevisiae,  
YBR097w] 1e-10  
[FUNCAT] 08.07 vesicular transport (golgi network, etc.) [S. cerevisiae, YBR097w]  
1e-10  
[FUNCAT] 30.08 organization of golgi [S. cerevisiae, YBR097w] 1e-10  
[FUNCAT] 06.04 protein targeting, sorting and translocation [S. cerevisiae, YBR097w]  
1e-10  
[FUNCAT] 10.04.99 other nutritional-response activities [S. cerevisiae, YJR059w]  
4e-09  
[FUNCAT] 01.06.10 regulation of lipid, fatty-acid and sterol biosynthesis [S.  
cerevisiae, YHR079c] 1e-07  
[FUNCAT] 30.07 organization of endoplasmic reticulum [S. cerevisiae, YHR079c]  
1e-07  
[FUNCAT] 08.19 cellular import [S. cerevisiae, YNL154c] 2e-04  
[BLOCKS] BL00415A Synapsins proteins  
[BLOCKS] BL00239B Receptor tyrosine kinase class II proteins  
[BLOCKS] BL00107A Protein kinases ATP-binding region proteins  
[SCOP] dlgol\_ 5.1.1.1.9 MAP kinase Erk2 [rat Rattus norvegicus 3e-78  
[SCOP] dlwfc\_ 5.1.1.1.8 MAP kinase p38 [human (Homo sapiens) 1e-81  
[SCOP] dikoa\_2 5.1.1.1.7 (1-350) Twitchin, kinase domain [Caenorhabditis 5e-89  
[SCOP] dikoba\_ 5.1.1.1.6 Twitchin, kinase domain [california sea har 5e-86  
[SCOP] dlphk\_ 5.1.1.1.5 gamma-subunit of glycogen phosphorylase kinase 3e-80  
[SCOP] dlirk\_ 5.1.1.2.4 insulin receptor [Human (Homo sapiens) 6e-70  
[SCOP] dlapme\_ 5.1.1.1.4 cAMP-dependent PK, catalytic subunit [mouse (Mu 1e-95  
[SCOP] dlfgka\_ 5.1.1.2.3 Fibroblast growth factor receptor 1 [human (Hom 7e-71  
[SCOP] dlydse\_ 5.1.1.1.3 cAMP-dependent PK, catalytic subunit [bovine (Bo 2e-96  
[SCOP] dlfmk\_3 5.1.1.2.2 (168-437) c-src tyrosine kinase [human (Hom 2e-72  
[SCOP] dlcdka\_ 5.1.1.1.2 cAMP-dependent PK, catalytic subunit [pig (Su 5e-97  
[SCOP] d2hckb3 5.1.1.2.1 (167-437) Haemopoietic cell kinase Hck [huma 2e-68  
[SCOP] dlcsn\_ 5.1.1.1.11 Casein kinase-1, CK1 [Schizosaccharomyces pombe 3e-53  
[SCOP] dljsua\_ 5.1.1.1.1 Cyclin-dependent PK [Human (Homo sapiens) 3e-78  
[SCOP] dickia\_ 5.1.1.1.10 Casein kinase-1, CK1 [rat (Rattus norvegicus) 1e-58  
[EC] 2.7.1.117 Myosin-light-chain kinase 3e-49  
[EC] 2.7.1.109 [Hydroxymethylglutaryl-CoA reductase (NADPH)] kinase 4e-78  
[EC] 2.7.1.138 Phosphorylase kinase 3e-41  
[EC] 2.7.1.137 Protein kinase 7e-45  
[EC] 2.7.1.123 Ca2+/calmodulin-dependent protein kinase 5e-42  
[EC] 2.7.1.128 [Acetyl-CoA carboxylase] kinase 4e-78  
[PIRKEW] phosphotransferase 3e-93  
[PIRKEW] nucleus 2e-74  
[PIRKEW] calcium 2e-40  
[PIRKEW] transferase 3e-33  
[PIRKEW] duplication 2e-32  
[PIRKEW] tandem repeat 7e-45  
[PIRKEW] phorbol ester binding 4e-33  
[PIRKEW] zinc 4e-33  
[PIRKEW] ion transport 1e-32  
[PIRKEW] cell cycle control 1e-45  
[PIRKEW] serine/threonine-specific protein kinase 2e-97  
[PIRKEW] oncogene 1e-34  
[PIRKEW] phospholipid binding 2e-32  
[PIRKEW] autophosphorylation 2e-74  
[PIRKEW] brain 6e-36  
[PIRKEW] heterotetramer 8e-38  
[PIRKEW] mitosis 1e-45  
[PIRKEW] polymer 5e-41  
[PIRKEW] magnesium 6e-80  
[PIRKEW] ATP 2e-97  
[PIRKEW] polyprotein 1e-34  
[PIRKEW] alternative initiators 2e-31  
[PIRKEW] phosphoprotein 2e-74  
[PIRKEW] apoptosis 8e-38  
[PIRKEW] cGMP binding 4e-33  
[PIRKEW] glycoprotein 3e-36  
[PIRKEW] skeletal muscle 8e-38  
[PIRKEW] protein kinase 2e-50  
[PIRKEW] testis 5e-41  
[PIRKEW] cAMP binding 8e-38  
[PIRKEW] transforming protein 4e-33  
[PIRKEW] purine nucleotide binding 7e-52  
[PIRKEW] calcium binding 7e-45  
[PIRKEW] alternative splicing 5e-42  
[PIRKEW] P-loop 7e-52  
[PIRKEW] lipoprotein 8e-38  
[PIRKEW] proto-oncogene 4e-33  
[PIRKEW] segmentation 1e-34  
[PIRKEW] core protein 1e-34



```

[PIRKW]      muscle 8e-38
[PIRKW]      myristylation 8e-38
[PIRKW]      EF hand 7e-45
[PIRKW]      cell division 3e-49
[PIRKW]      homodimer 1e-32
[PIRKW]      calmodulin binding 5e-42
[SUPFAM]     ribosomal protein S6 kinase II 1e-34
[SUPFAM]     calcium-dependent protein kinase 7e-45
[SUPFAM]     AMP-activated protein kinase 6e-80
[SUPFAM]     protein kinase akt 3e-36
[SUPFAM]     protein kinase SPK1 7e-41
[SUPFAM]     unassigned Ser/Thr or Tyr-specific protein kinases 8e-99
[SUPFAM]     Ca2+/calmodulin-dependent protein kinase 5e-42
[SUPFAM]     calmodulin repeat homology 7e-45
[SUPFAM]     cAMP receptor protein cyclic nucleotide-binding domain homology 3e-33
[SUPFAM]     protein kinase DUN1 6e-36
[SUPFAM]     protein kinase C zeta 4e-33
[SUPFAM]     Dictyostelium cAMP-dependent protein kinase catalytic chain 2e-34
[SUPFAM]     death-associated protein kinase 8e-38
[SUPFAM]     pleckstrin repeat homology 3e-36
[SUPFAM]     ankyrin repeat homology 8e-38
[SUPFAM]     protein kinase homology 8e-99
[SUPFAM]     Ca2+/calmodulin-dependent protein kinase II 6e-38
[SUPFAM]     protein kinase C zinc-binding repeat homology 4e-33
[SUPFAM]     protein kinase C delta 2e-32
[SUPFAM]     cGMP-dependent protein kinase 3e-33
[SUPFAM]     protein kinase cdrl 1e-45
[SUPFAM]     kinase-related transforming protein 2e-50
[SUPFAM]     Ca2+/calmodulin-dependent protein kinase I 8e-42
[SUPFAM]     kinase interaction domain homology 7e-41
[SUPFAM]     gag-akt polyprotein 1e-34
[PROSITE]    PROTEIN_KINASE_ATP      1
[PROSITE]    MYRISTYL                 3
[PROSITE]    AMIDATION                 2
[PROSITE]    CAMP_PHOSPHO_SITE         4
[PROSITE]    CK2_PHOSPHO_SITE          15
[PROSITE]    TYR_PHOSPHO_SITE           2
[PROSITE]    PKC_PHOSPHO_SITE           10
[PROSITE]    ASN_GLYCOSYLATION          2
[PROSITE]    PROTEIN_KINASE_ST         1
[PFAM]       Eukaryotic protein kinase domain
[KW]         Irregular
[KW]         3D
[KW]         LOW_COMPLEXITY      12.31 %

```

```

SEQ      MVMADGPRHLQRPVRVGFYDIEGLGKGNFAVVKLGRHRITKTEVAIKIIDSQQLDAVN
SEG      .....
lctpe    .....EEECTTTEEEEEETTTTEEEEEEEHHHHHHHC

SEQ      LEKIYREVQIMKMLDHPHIKLYQVMTKSMYLVTEYAKNGEIFYDLANHGRNLNESEAR
SEG      .....
lctpe    HHHHHHHHHHHHCCCTTTBCCEEEEEETTEEEEEECTTTTBHHHHHHHHHCCCHHHH

SEQ      RKFQILSAVDYCHGRKIVHRDLKAENLLDNNMNKIIDPFGNFFKSGELLATWCGSP
SEG      .....
lctpe    HHHHHHHHHHHHHHCCCECCCCGGGEETTTTCEEEECTTTTEETT-TTBC-CCCCCG

SEQ      PYAAPEVFEGQQYEGPQLDIWSMGVVLVLCGALPFDGPTLPILRQRVLEGRFRIPIYFM
SEG      .....
lctpe    GGCCHHHHHCCBC-HHHHHHHHHHHHHHCCCTTTTTHHHHHHHHHHCCCTTTT

SEQ      SEDCEHLIRMLVLDPSKRLTIAQIKHKWMLIEVPVQRPVLPQEQENEPSIGEFNEQV
SEG      .....
lctpe    CHHHHHHHHTTTTGGGTTTHHHHHHCGG.....

SEQ      LRLMHSIGIDQQKTIESLQNSYNHFAAIYFLLVERLKSRRSFPVEQRLDGRQRRPSTI
SEG      .....
lctpe    .....

SEQ      AEQTVAKAQTVGLPVTMHSNMLRLSALLPQASNVEAFSPASGCQAEAAFMEEECVDT
SEG      .....
lctpe    .....

SEQ      PKVNGCLLDPPVPLVRKGCQSLPSNMETSIDEGLETEGEAEEDPAHAFAEAFQSTRSGQ
SEG      .....
lctpe    .....XXXXXXXXXX.....

SEQ      RRHTLSEVTNQLVVMGAGKIFSMNDSPLSDVDSEYDMGSVQVDLNFLEDNPSLKDIML
SEG      .....
lctpe    .....

```

```

SEQ      ANQPSRMTSPFISLRPTNPAMQALSSQKREVHNRSPVSFREGRRASDTSLTGIVAFRQ
SEG      .....
lctpe    .....

SEQ      HLQNLARTKGILELNKVLLEYQIGPEADPNLAPAPQLQDLASSCPQEEVSQQQESVST
SEG      .....XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
lctpe    .....

SEQ      LPASVHPQLSPROSLETQYLQHLRQKPSLLSKAQNTCQLYCKEPPRSLEQQLQEHRLQK
SEG      .....XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
lctpe    .....

SEQ      RLFLQKSQLQAYFNQMQUIAESSYPQPSQQLPLPRQETPPPSQQAPPFSLTQPLSPVLEP
SEG      .....XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
lctpe    .....

SEQ      SSEQMYSFPLSQYQEMQLQPLPSTSGPRAAPPLPTQLQQQPPPPPPPPPPRQPCAAPA
SEG      .....XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
lctpe    .....

SEQ      PLQFSYQTCELPSAASPAPDYPTPCQYPVQGAQQSDLTGPDGPRSPGLQEPSSSYDPLAL
SEG      .....
lctpe    .....

SEQ      SELPGLFDCMLDAVDOPQHNGYVLVN
SEG      .....
lctpe    .....

```

## Prosites for DKFZphtes3\_15k11.1

PS00001	115->119	ASN_GLYCOSYLATION	PDOC00001
PS00001	320->324	ASN_GLYCOSYLATION	PDOC00001
PS00004	258->262	CAMP_PHOSPHO_SITE	PDOC00004
PS00004	355->359	CAMP_PHOSPHO_SITE	PDOC00004
PS00004	481->485	CAMP_PHOSPHO_SITE	PDOC00004
PS00004	584->588	CAMP_PHOSPHO_SITE	PDOC00004
PS00005	257->260	PKC_PHOSPHO_SITE	PDOC00005
PS00005	339->342	PKC_PHOSPHO_SITE	PDOC00005
PS00005	420->423	PKC_PHOSPHO_SITE	PDOC00005
PS00005	475->478	PKC_PHOSPHO_SITE	PDOC00005
PS00005	534->537	PKC_PHOSPHO_SITE	PDOC00005
PS00005	545->548	PKC_PHOSPHO_SITE	PDOC00005
PS00005	554->557	PKC_PHOSPHO_SITE	PDOC00005
PS00005	567->570	PKC_PHOSPHO_SITE	PDOC00005
PS00005	579->582	PKC_PHOSPHO_SITE	PDOC00005
PS00005	670->673	PKC_PHOSPHO_SITE	PDOC00005
PS00006	42->46	CK2_PHOSPHO_SITE	PDOC00006
PS00006	54->58	CK2_PHOSPHO_SITE	PDOC00006
PS00006	128->132	CK2_PHOSPHO_SITE	PDOC00006
PS00006	292->296	CK2_PHOSPHO_SITE	PDOC00006
PS00006	359->363	CK2_PHOSPHO_SITE	PDOC00006
PS00006	394->398	CK2_PHOSPHO_SITE	PDOC00006
PS00006	450->454	CK2_PHOSPHO_SITE	PDOC00006
PS00006	458->462	CK2_PHOSPHO_SITE	PDOC00006
PS00006	484->488	CK2_PHOSPHO_SITE	PDOC00006
PS00006	503->507	CK2_PHOSPHO_SITE	PDOC00006
PS00006	515->519	CK2_PHOSPHO_SITE	PDOC00006
PS00006	534->538	CK2_PHOSPHO_SITE	PDOC00006
PS00006	579->583	CK2_PHOSPHO_SITE	PDOC00006
PS00006	878->882	CK2_PHOSPHO_SITE	PDOC00006
PS00006	893->897	CK2_PHOSPHO_SITE	PDOC00006
PS00007	672->680	TYR_PHOSPHO_SITE	PDOC00007
PS00007	100->108	TYR_PHOSPHO_SITE	PDOC00007
PS00008	372->378	MYRISTYL	PDOC00008
PS00008	871->877	MYRISTYL	PDOC00008
PS00008	905->911	MYRISTYL	PDOC00008
PS00009	134->138	AMIDATION	PDOC00009
PS00009	582->586	AMIDATION	PDOC00009
PS00107	26->50	PROTEIN_KINASE_ATP	PDOC00100
PS00108	138->151	PROTEIN_KINASE_ST	PDOC00100

## Pfam for DKFZphtes3\_15k11.1

HMM\_NAME      Eukaryotic protein kinase domain

HMM		*YeigRiIGeGsFGtVYkCiWr.TGeIVAiKIIkkrsms.....FlREI	
Query	20	YDIEGTLGKGNFAVVKLGRHRITKTEVAIKIIDKSQLDAVNLEKIYREV	68
HMM		qIMRrLnHPNIIRFYDwFedddDHIYMIMEYMeGGDLFDYIrrngpMsEw	
Query	69	QIM++L+HP+II++Y ++E +++ +Y+++EY+ +G++FDY+ ++G+++E	117
HMM		eIrfIMyQILrGMeYLHSMgIIHRDLKPENILIdengqIKicDFGLARqM	
Query	118	EARRKFWQILSAVDYCHGRKIVHRDLKAENLLLDNNMNIKIADFGFNGFF	167
HMM		nnYerMttfCGTPWYMMAPeVIimg.nyYttkVDMWSFGCILWEMMTGep	
Query	168	+++E++ T CG+P+Y APEV +G +Y +++ D+WS+G++L+ +++G +	215
HMM		PFyddnMemImriIqrfrrpfWpnCSeElyDFMrcWnyDPekRPTFrQI	
Query	216	PF++ ++ + + +++ R++++ +SE++ +++R+++ +DP+KR+T+ QI	265
HMM		LnHPWF*	
Query	266	+H W+ KEHKWM 271	

DKFZphtes3\_17f10

group: testes derived

DKFZphtes3\_15j18 encodes a novel 710 amino acid protein with weak similarity to neurofilament proteins.

No informative BLAST results; No predictive prosite, pfam or SCOP motif.

The new protein can find application in studying the expression profile of testis-specific genes.

similarity to neurofilament proteins

Sequenced by GBF

Locus: unknown

Insert length: 2533 bp

Poly A stretch at pos. 2507, no polyadenylation signal found

```
1 CTTCAGTTCA ACTAAAAATG GACAGATCTC AGCAGACCAG CCGTACAGGA
51 TACTGGACCA TGATGAACAT CCCCCCTGTA GAAAAAGTGG ACAAGGAACA
101 ACAGACATAC TTTAGTGAAT CAGAAATAGT GGTATTTC C AGGCCAGATA
151 GTTCTTCTAC AAAGTCAAAG GAAGATGCC C TGAAACATAA ATCGTCGGGA
201 AAGATTTTGT CTAGTGAAAC CCCTGAATTT CAACCAAGCAA CAACACGCAA
251 TGAAGAAATT GGGCAGAAAA ATATCAGCAG AACTTCATTT ACTCAGGAGA
301 CTAAAAAAGG TCCCCAGTA CTTTGAAGG ATGAGCTTAG GGAAGAAGTA
351 ACTGTACCTG TTGTACAAGA AGGTTCTGCT GTTAAAAAAG TGGCTTCTGC
401 TGAATATAGG CCTCCATCAA CAGAAAAATT CCCAGCTAAA ATACAGCCTC
451 CATTAGTTGA AGAGGCCACT GCTAAAGCGG AGCCCAGACC TGCTGAAGAG
501 ACCCATGTCC AAGTACAGCC ATCAACTGAA GAGACTCCTG ATGCTGAGGC
551 AGCCACTGCA GTTGCAGAGA ATTCTGTTAA AGTTCAGCCT CCACCTGCTG
601 AAGAGGCCCC TTTAGTGGAG TTTCTGCTG AAATTCAGCC TCCATCAGCT
651 GAAGAGTCTC CTTCTGTAGA GCTTCTGGCT GAAATTCCTG CTCCATCAGC
701 TGAAGAGTCC CTTTCAGAA AGCCTCCTCG TGAAATTCCT CCTCCACCAG
751 CTGAAAAAGT TCCTTCAGTA GAGCTTCTTG GTGAAATTCG GTCTCCCTCA
801 GCACAAAAGG CTCCCATTTA AGTACAGCCT TTACAGCTGT AGGGCCGCTT
851 TGAAGAGGCC CCAGCTAAAG TAGAGCCTCC CACTGTTGAA GAGACCCCTT
901 CTGAAGTTCA GCCTCTATTA CCTGAAGAGG CTCTAGAGA AGAGGCTCGA
951 GAACTTCAGC TTTCAACAGC TATGGAGACC CCGCAGAAAG AGGCTCCTAC
1001 TGAATTTTCA TCTCCATTAC CTAAGAGAGC CACTGCAGAA GAGGCCTCTG
1051 CTGAAATTCA GCTTCTAGCA GCTACGGAGC CTCTGCAGAA TGAAATCTCT
1101 GCCGAAGCTC GGTCTCCACT ATCTGAGGAG ACTTCTGCAG AAGAGGCTCA
1151 TGCTGAAGTT CAATCTCCAT TAGCTGAAGA GACCACTGCA GAAGAGGCTT
1201 CTGCTGAAAT TCAGCTCTTA GCAGCTATAG AGGCTCCTGC AGATGAAACT
1251 CCTGCTGAAG CTCAGTCTCC ACTATCTGAG GAGACTTCTG CAGAAGAGGC
1301 TCCTGCTGAA GTTCAGTCTC CATCAGCTAA GGGAGTTTCT ATAGAAGAGG
1351 CCCCTCTTGA GCTTCAGCCT CCATCAGGTG AAGAGACCAC TGCAGAAGAG
1401 GCCTCTGCTG CAATTACAGT TCTAGCAGCT ACAGAGGCTT TGCAGAAGA
1451 GGCTCCTGCT GAAGTTCAGC CTCCACCAGC TGAGGAGGCC CCCGCTGAAG
1501 TTACGCTCC ACCAGCTGAG GAGGCCCCCG CTGAAGTTCA GCCTCCACCA
1551 GCTGAGGAGG CCCCCTGTA AGTTTCAGCT CCACCACTG AGGAGGCCCC
1601 CGCTGAAGTT CAGCTCCAC CAGCTGAGGA GGGCCCCGCT GAAGTTCAGC
1651 CTCCACCAGC TGAGGAGGCC CCCTCTGAAG TTCAGCTCC ACCAGCTGAG
1701 GAGGCCCTG CTGAAGTTCA GTCTCTACCA GCTGAGGAGA CTCCTATAGA
1751 AGAGACCCCT GCTGCAGTAC ACTCTCCCCC AGCTGATGAT GTCCCTGCAG
1801 AAGAGGCCCT CGTTGACAAA CATTCCCCAC CAGCTGATTT GCTTCTGACT
1851 GAGGAGTTTC CTATAGGAGA GGCCTCTGCT GAAGTTTCAC TCCACCATC
1901 TGAACAAACC CCTGAAGATG AGGCTCTGGT AGAAGATGTG TCTACAGAAT
1951 TTCAGTCACC GCAGGTGGCA GGAATTCAG CAGTAAAAAT AGGATCGGTT
2001 GTTTTGGAAG GTGAAGCAAA ATTTGAAGAG GTTTCAAAA TCAATTCTGT
2051 CCTTAAAGAT TTGTCTAATA CCAATGATGG ACAGGCTCCC ACTCTTGAAA
2101 TAGAAAGTGT TTTTCATATA GAAATTAAC AACGTCCTCC TGAAGTGTAG
2151 TCAGGTTGTA CCTAAGCTAG CAATCAGAAG CTACATGGTT TTGGAAGAAC
2201 ATACTTTAGA AAAGGTTGGG CAGCAGGAAG TAGCTTTGTC AATAAGGCAA
2251 ATTAAGGGG ACCCAAGAC TTGGAATACA GTTTGAAGAA TGAACAATAA
2301 AAAGTGTAGC AGCAATAAAT TACTTGTGTT AATTTCATTC AAATTTATGG
2351 CATGAAAAAT ACCTATTTTG AAAGTAAGTT TATAATTGAA AAAAATTGCT
2401 TAAAAATACC TTCTACAGT AAAGTGTGTT ACACAGTAA AGTTTAATCT
2451 GCAGCCATCT TTTCTGTCT TTGCCTTCCC TTTATAAGTA AATATAGTTT
2501 CTAGTGGAAA AAAAAAAAAA AAAAAAAAAA AAA
```

BLAST Results

-----

No BLAST result

## Medline entries

No Medline entry

## Peptide information for frame 3

ORF from 18 bp to 2147 bp; peptide length: 710  
Category: similarity to known protein  
Classification: unclassified

1 MDRSQTSRT GYWTMMNIPP VEKVDKEQQT YFSESEIVVI SRPDSSTKS  
51 KEDALKHKSS GKIFASEHPE FQPATNSNEE IGQKNISRTS FTQETKKGPP  
101 VLLEDELREE VTPVVQEGS AVKKVASAEI EPPSTEKFPK KIQPPPLVEEA  
151 TAKAEPRPAE ETHVQVPST EETPDAAEAT AVAENSVKQV PPPAEAPLV  
201 EFPAEIQPPS AEESPSVELL AEILPPSAEE SPSEEPFAEI LPPPAEKSPS  
251 VELLGEIRSP SAQKAPIEVQ PLPAEGALEE APAKVEPPTV EETLAEVQPL  
301 LPEEAPREEA RELQLSTAME TPAEEAPTFF QSPLPKETTA EEASAEIQLL  
351 AATEPPADET PAEARSPLSE ETSAAEAHAE VQSPLAEETT EEASAEIQL  
401 LAIEAPADE TPAEAQSPLS EETSAAEAPA EVQSPSAGKV SIEEAPLELQ  
451 PPSGEETTAE EASAAIQLLA ATEASAEAP AEVQPPPAEE APAEVQPPPA  
501 EEPAEVQPP PAEEAPAEVQ PPPAEAPAE VQPPPAEAP AEVQPPPAEE  
551 APSEVQPPPA EEPAEVQSL PAEETPIET LAHVHSPPAD DVPAAEASVD  
601 KHSPPADLLL TEEFPIGEAS AEVSPPPEEQ TPEDEALVEN VSTEFQSPQV  
651 AGIPAVKLGS VVLEGEAKFE EVSKINSVLK DLSNTNDGQA PTEIESVVFH  
701 IELKQRPPEL

## BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphtes3\_17f10, frame 3

PIR:A37221 neurofilament triplet H protein - rat, N = 1, Score = 480, P  
= 7.4e-43

TREMBL:RNNFLH\_1 Rat heavy neurofilament subunit (NF-H) mRNA, 3' end., N  
= 1, Score = 475, P = 1e-42

>PIR:A37221 neurofilament triplet H protein - rat  
Length = 1,072

## HSPs:

Score = 480 (72.0 bits), Expect = 7.4e-43, P = 7.4e-43  
Identities = 185/622 (29%), Positives = 320/622 (51%)

Query: 33 SESEIVVISRPDSSTKSKEDALKHKSSGKIFASEHPEFQPATNSNEEIGQKNISRTSFT 92  
SE +I V+ + + + +E + + + ++ E E Q E G + + TS  
Sbjct: 436 SEEKIKVVEKSEKETVIVEEQTEEIQTVEVTEEDKEAQGEEEEAEEGGEEAATTSPP 495

Query: 93 QETKKGPPVLLEDELREEVTPVVQEGSAVKKVASAEIEPPSTEKFPKIQPPPLVEEATA 152  
E P + ++EE P + A K + AE + P+ K PA+++ P ++ A  
Sbjct: 496 AEEAASPEKETKSPVKEEAKSPAEEKSPAEEK-SPAEEKSPAEEKSPAEEKSPA 554

Query: 153 KAEPRPAEETHVQVPSTTEETPDAAEATAVAENSVKQPPPAEEAP-LVEFPAEIQPPSA 211  
+A+ PAE V+ P+T ++P + A A++ +V+ P ++P + PAE + P+  
Sbjct: 555 EAKS-PAE---VK-SPATVKSPAEEKSPAEEKSPAEEKSPAATVKSPGEAKSPAEEKSPA 609

Query: 212 EESP-SVELLAELPPSAEESPSE-EPPAEILPPPAEKSPS-VELLGEIRSPSAQKAPIE 268  
+SP + AE P++ +SP E + PAE P KSP+ V+ E +SP+ K+P+  
Sbjct: 610 VKSPVEAKSPAEEKSPASVKSPGEAKSPAEEKSPAEEKSPAATVKSPVEAKSPAEEKSPVT 669

Query: 269 VQPLPAEGALEEAPAKVEPPTVEETLAEVQPLLPPEAPREEARELQLSTAMETPAE-EAP 327  
V+ PAE ++P +V+ P ++ +E + ++P E A+ ++PAE ++P  
Sbjct: 670 VKS-PAEA---KSPVEKSPASVKSPSEAKSPAGAKSPA-EAKS---PVVAKSPAEEKSP 721

Query: 328 TEFQSPLPKETTAEEASAEIQLLAATEPPAD-ETPAEARSPLSEETSAAEAHAEVQS--- 383  
E + P ++ AE S A + PA+ ++PAE+SP+ E S E+A + V+  
Sbjct: 722 AEAKPPEAKSPAEEKSP-----AEAKSPAEEKSPAEEKSPV-EVKSPEKAKSPVKEGAK 775

Query: 384 PLAEETTAEEASAEIQLLAIEAPAD-ETPAEAQSPLSEET-SAEAPAE-EVQSPSAGKV 440

LAE + E+A + ++ I+ PA+ ++P +A+SP+ EE S E+A +V+SP AK  
Sbjct: 776 SLAEAKSPEKAKSPVK--EEIKPPAEVKSPEKAKSPMKKEAKSPEKAKTLDVKSPEAKTP 833

Query: 441 SIEEA--PLELQPPSGEETTA--EASAAIQLLAATEASA---EEAPAEVQPPPAEEAPAE 494  
+ EEA P +++ P ++ A EEA + + TE A EE + V+ A+E P +  
Sbjct: 834 AKKEAKRPADIRSPQVKSPEAKSPEKEETRTKVPKKEEVKSPVEEVKAKEPPKK 893

Query: 495 VQPPPAEEAP-AEVQPPPAEEAPAEVQPPPAEEAPAEVQPPPAEEAPAEVQPPPAEEAPS 553  
V+ P EV+ +EAP E Q P AEE + P +++P E + EEA  
Sbjct: 894 VEEETPATPKTEVKESKKDEAPKEAQPKAEKEPLETEKP--KDSPEAKK---EEAKE 948

Query: 554 EVQPPPAEEAPAEV---QSLP---AEETPIEETL--AAVHSPADDVPAEEASVD-KHS 603  
+ P EE PA++ ' ++ P AE+ +E + P ++VPA D K  
Sbjct: 949 KAAAPEEETPAKLGVKKEAKPKAEADAKAKEPSKPSEKEPKKEEVPAAPKKKDTKEE 1008

Query: 604 PPADLLTEEFFIGEASAEVSP--PSEQT-PEDEALVENVSTEFQSPQ 649  
+ EE P +A A+ P E + P+ E ++ ST+ + Q  
Sbjct: 1009 KTESKKPEEKPMQAKAKEEDKGLPQEPSPKPKTEKAESSSTDQKDSQ 1057

Score = 473 (71.0 bits), Expect = 4.8e-42, P = 4.8e-42  
Identities = 184/628 (29%), Positives = 310/628 (49%)

Query: 18 IPPVEKVDKEQTYFSESEIVVISRP---DSSSTKSKEDALKHKSSGKIFASEHPEFQPA 74  
I VEK +KE ++E + ++ + E+ + + G+ A+ P + A  
Sbjct: 440 IKVVEKSEKETVIVEEQTEEIQTVEETVEEDKEAQEEEEAEEGGEAAATSPPAEEA 499

Query: 75 TNSNEEIGQKNISRTSFTQETKKGPPVLEDELREEVTVPVVQEGSAVKKVASAEIEPPS 134  
+ +E + + + K P E + E P + A K + AE + P+  
Sbjct: 500 ASPEKET-KSPVKEAKSPAEAKSPA---EAKSPAEEKSPAEEVKSPEAVK-SPAEEKSPA 554

Query: 135 TEKFPKIQPPLVEEATAKAEPRAETHVQVQ-PSTEETPDAAEATAVAENSVKVQPPP 193  
K PA+++ P ++ A+A+ ++ +V+ P+T ++P + A A++ +V+ P  
Sbjct: 555 EAKSPAEEVKS PATVKS PAEAKSPAEEKSPAEEVKS PATVKS PGEAKSPAEEKSPAEEVKS P 614

Query: 194 AEEAPL-VEFPAEIQPPSAEESPS-VELLAELPPSAEESPS-EPPAEILPPPAEKSPS 250  
++P + PA ++ P +SP+ + AE+ P+ +SP E + PAE+ P KSP+  
Sbjct: 615 EAKSPAEEKSPASVKS PGEAKSPAEEKSPAEEVKS PATVKS PVEAKSPAEEKSPAEEVKS P 674

Query: 251 -VELLGEIRSPSAQKAPIEVQ-PLPAEGALE-EAPAKVEPPTVEETLAEVQPLLPEEAPR 307  
+ E++SP++ K+P E + P A+ E ++P + P ++ AE + P ++P  
Sbjct: 675 EAKSPVEVKS PASVKS PSEAKSPAGAKSPAEEKSPVAKSPAEEKSPAEEKSPAEEKSPA 734

Query: 308 EEARELQLSTAME--TPAE-EAPTEFQSP----LP-KE---TTAEASAEIQLLAATE-- 354  
E + + E +PAE ++P E +SP P KE + AE S E E  
Sbjct: 735 EAKSPAEEKSPAEEKSPAEEKSPAEEKSPAEEKSPAEEKSPAEEKSPAEEKSPAEEKSPA 794

Query: 355 -PPAD-ETPAEARSPLSEET-SAEAAH-EVQSPLAETTTAEAS--AEIQLLAATEAPA 408  
PPA+ ++P +A+SP+ EE S E+A +V+SP A+ EEA A+I+ ++P  
Sbjct: 795 KPPAEVKSPEKAKSPMKKEAKSPEKAKTLDVKSPEAKTPAKEEAKRPADIRSPQVKS 854

Query: 409 DETPAEAQSPLEETSAAE-APA--EVQSPSAGVSIIEAPLELQPPSGEETTAEEASAA 465  
E EA+SP EET E+ AP EV+SP +EE + +PP E EE + A  
Sbjct: 855 KE---EAKSPEKEETRTKVPKKEEVKSP-----VEEVKAK-EPPKKVE---EETPA 901

Query: 466 IQLLAATEASAEAPAEVQPPPAEEAPAEVQPPPAEEAPAEVQPPPAEEAPAEVQPPPAE 525  
E+ +EAP E Q P AEE + P +++P E + A+E A P E  
Sbjct: 902 TPKEVKSKEKKDEAPKEAQPKAEKEPLETEKP--KDSPEAKKEEAKKAAA---PEE 956

Query: 526 EAPAEV---QPPPAEEAPAEVQPPPAEEAPSEVQPPPAEEAPAEVQSLPAEETPIEETL 581  
E PA++ + P E+A P++ PSE + P EE PA + +E E+  
Sbjct: 957 ETPAKLGVKKEAKPKAEADAKAKEPSK--PSEKEPKKEEVPAAPKKKDTKEETTESK 1014

Query: 582 AAVHSPADDVPAEEASVDKHSPPADLL-LTEEFPIGEASAEVSPPPSEQTPEDEA 636  
P EE DK P TE+ ++ + PSE+ PED+A  
Sbjct: 1015 KPEEKPMQAKAKEE---DKGLPQEPSPKPKTEKAESSSTDQKDSQSEKAPEDKA 1067

Score = 421 (63.2 bits), Expect = 3.7e-36, P = 3.7e-36  
Identities = 162/540 (30%), Positives = 275/540 (50%)

Query: 135 TEKFPKIQPPLVEEATAKAEPRAETHVQVQSTEETPDAAEATAVAENSVKV 189  
TE P KI P + K+E + +E+ V V+ TEE E T E +  
Sbjct: 419 TEGLP-KI-PSMSTHIKVKSEKIKVVEKSEKETVIVEEQTEEIQTVEETVE--EEDKEA 474

Query: 190 QPPPAEEAPLVEFPAEIQPPSAEESPSVELLAELPPSAEE--SPSE-EPPAEILPPPAE 246  
Q EEA A P AEE+ S E E P EE SP+E + PAE P  
Sbjct: 475 QGEEEEAEEGGEAAATSPPAEEAASPE--KETKSPVKEAKSPAEEKSPAEEKSPA 532

Query: 247 KSPSVELLGEIRSPSAQKAPIEVQPLPAEGALEEAPAKVEPPTVEETLAEVQPLLPEEAP 306  
KSP+ E++SP+ K+P E + PAE ++PA+V+ P ++ AE + +P  
Sbjct: 533 KSPA-----EVKSPAEEVKS PAEAKS-PAEA---KSPAEEVKS PATVKS PAEAKSPAEEKSPA 583

Query: 307 REEARELQLSTAME--TPAE-EAPTEFQSPLPKETTAEAS-AEIQLLAATEPPAD-ETP 361  
 E + + E +PAE ++P E +SP+ ++ AE S A + + + PA+ ++P  
 Sbjct: 584 AEVKSPTATVKSPEAKSPAEAKSPAEVKSPEAKSPAEAKSPASVKSPEAKSPAEAKSP 643

Query: 362 AEARSPLSEETSAE-EAHAEVQSPLAEETTAEEASAEIQLLAAIEAPAD-ETPAEQSPL 419  
 AE +SP + ++ E ++ AEV+SP+ ++ AE A + ++ +++PA ++P+EA+SP  
 Sbjct: 644 AEVKSPTATVKSPEAKSPAEVKSPTVKSPEAE-AKSPVE----VKSPASVKSPEAKSP- 697

Query: 420 SEETSAAEAPAEVQSPS-AKGVSIIEAPLELQPPSGEETTAEASAAIQLLAATEASAE 478  
 + ++PAE +SP AK + ++P E +PP+ ++ AE S A + A + A A+  
 Sbjct: 698 ----AGAKSPAEAKSPVAKSPAEAKSPAEAKPPAEAKSPAEAKSPAE---AKSPAEAK- 749

Query: 479 APAEVQPPPAEEAPAEVQPPPAEEAP--AEVQPPPAEEAPA--EVQPPPAEEAPAEVQPP 534  
 +PAE + P ++P + + P E A AE + P ++P E+PP ++P + + P  
 Sbjct: 750 SPAEAKSPVEVKSPEAKSPVKEGAKSLAEAKSPKAKSPVKEEIKPPAEVKSPEAKSP 809

Query: 535 PAEEAPAEVQPPPAEEAPSEVQPPPAEEA--PAEVQSLPAEETPIETLAHVHSPADDV 592  
 EEA + + + E + P EEA PA+++S ++P E SP ++  
 Sbjct: 810 MKEEAKSPEKAKTLDVKSPEAKTPAEKRAKPADIRSPQVKSPEKE---AKSPEKEET 866

Query: 593 PAEEASVDKHS--PPADLLTTEFFIGEASAEVSPPPSEOTPEDEALVENVSTEFOSQOV 650  
 E+ + K P + + +E P + E P + +T E Q P+  
 Sbjct: 867 RTEKVAPKKEEVKSPEVKEVKAKEPP--KKVEEEKTPATPKTEVKESKKDEAPKEAQKPKA 924

Query: 651 AGIPAVKLGSVVLEGEAKFEEVSK 674  
 + GEAK EE +  
 Sbjct: 925 EEKEPLTEKPKDSPGEAKKEEAK 948

Score = 406 (60.9 bits), Expect = 1.7e-34, P = 1.7e-34  
 Identities = 123/390 (31%), Positives = 213/390 (54%)

Query: 308 EEARELQLSTAMETPAEEAPTEFQSPLPKETTAEASAEIQLLAATEPPADETPA---EA 364  
 E+ E+Q++ E EE E Q +E AEE E A T PPA+E + E  
 Sbjct: 455 EQTEEIQVT---EEVTEEDKEAQGE---EEEEAEEGGEEA---ATTSPPAEAAPEKET 506

Query: 365 RSPLSEETSAAEHAHVQSPLAEETTAEEAS-AEIQLLAAIEAPAD-ETPAEQSPLSEE 422  
 +SP+ EE + AE +SP ++ AE S AE++ A +++PA+ ++PAEA+SP +  
 Sbjct: 507 KSPVKEEAKSP---AEAKSPAEAKSPAEAKSPAEVKSPEAKSPAEAKSPAEAKSPAEV 563

Query: 423 TSAE-EAPAEVQSPS-AKGVSIIEAPLELQPPSGEETTAEASAAIQLLAATEASAEAP 480  
 + A ++PAE +SP+ AK + ++P ++ P GE E A + ++ EA + ++P  
 Sbjct: 564 SPATVKSPEAKSPAEAKSPAEVKSPTATVKSPEAKSPAEAKSPAEVKSPEV---KSP 619

Query: 481 AEVQPPPAEEAPAEVQPPPAEEAPAEVQPPPAEEAPAEVQPPPAEEAPAEVQPPPAEEAP 540  
 AE + P + ++P E + P ++PAEV+ P ++P E + P ++P V+ P ++P  
 Sbjct: 620 AEAKSPASVKSPEAKSPAEAKSPAEVKSPTATVKSPEAKSPAEVKSPTVKSPEAKSP 679

Query: 541 AEVQPPPAEEAPSEVQPPPAEEAPAEVQSLPAEETPIETLAHVHSPAD-DVPAEEASV 599  
 EV+ P + ++PSE + P ++PAE +S ++P E A PPA+ PAE S  
 Sbjct: 680 VEVKSASVKSPEAKSPAGAKSPAEAKSPVAKSPAEAKSPAEAKPPAEAKSPAEAKSP 739

Query: 600 DKHSPADLLTTEFFIGEASAEVSPPPSEOTPEDEALVENVSTEFOSQOVAGIPAVKLG 659  
 + PA+ E ++ EV P ++P E +++ E +SP+ A P VK  
 Sbjct: 740 AEAKSPAEAKSPAE---AKSPVEVKSPEAKSPVKEG-AKSLA-EAKSPEAKSP-VK-E 792

Query: 660 SVVLEGEAKFEEVSKINSVLKDLSTNDGQAPTLIES 697  
 + E K E +K S +K+ + + +A TL+++S  
 Sbjct: 793 EIKPPAEVKSPEKAK--SPMKEAKSPE-KAKTLDVKS 827

Score = 255 (38.3 bits), Expect = 5.5e-18, P = 5.5e-18  
 Identities = 124/420 (29%), Positives = 199/420 (47%)

Query: 252 ELLGEIRSPSAQKAPIEVQPLPA-----EGALEEAPAKVEPPTVEETLAHVQPLLPPEAP 306  
 ELLG+I+ A +A + + A AL E A++E TV+ TL +  
 Sbjct: 236 ELLGQIQCGGAQAQAQAEARDALKCDVTSALREIRAQLEGHTVQSTLQSEWFRVRLDR 295

Query: 307 REEARELQLSTAMETPAEEAPTEFQSPLPKETTAEASAEIQLLAATEPPADETPAEARS 366  
 EA ++ + AM + EE TE++ L TT E++ L +T+ + +E  
 Sbjct: 296 LSEAAKVN-TDAMRSAQEET-TEYRRQLQARTT-----ELEALKSTKESLERQRSELED 347

Query: 367 PLSEE-TSAEEAHAHVQSPLAEETTAEEASA--EIQLLAAIEAPAD-ETPAEQSPLSEE 422  
 + S ++A ++ + L T E A+ E Q L ++ D E A + EE  
 Sbjct: 348 RHQVDMASYQDAIQQLDNEL-RNTKWEAAQLREYQDLLNVKMLDIEIAAYRKLEGE 406

Query: 423 TSAEEAPAEV-----QSPS-AKGVSI-EAPLELQPPSGEETT-AEEASAAIQLLA-A 471  
 P+ + PS + + ++ E +++ S +ET EE + IQ+  
 Sbjct: 407 CRIGFGSPSFLTEGLPKIPSYSTHIKVKSEEKIKVVEKSEKTVIVEEQTEEIQTVEEV 466

Query: 472 TEASAEAPAEVQPPPAEEAPAEVQPP--PPAEAPAEVQPPPAEEA--PAEVQPPPA 524  
 TE +EA E + AEE E PAEEA + E + P EEA PAE + P  
 Sbjct: 467 TEEEDKEAQGE-EEEEAEEGGEEAATTSPPAEAAASPEKETKSPVKEAKSPAEAKSPAE 525

```

Query: 525 EEAPAEVQPPPAEEAAEPAEVQPPPAEEAPSEVQPPPAEEAAEPAEVQSLPAE-ETPIE-ETLA 582
      ++PAE + P ++PAEV + P ++P+E + P ++PA V+S PAE ++P E ++ A
Sbjct: 526 AKSPAEKSPAEVKSPEAKSPAEAKSPAEVKS PATVKS-PAEAKSPAEAKSP 584

Query: 583 AVHSPPADDPVPAEASVDKHSPPADLLLTTEEFIGEASAEVSPPPSEQT-P-EDEALVENV 641
      V SP P S E S + PA++ E ++ AE P S ++P E ++ E
Sbjct: 585 EVKSPATVKSPEGAEKSPAEAKSPAEVKS PVE--AKSPAEAKSPASVKSPGEAKSPAEAK 641

Query: 642 S-TEFQSPQVAGIP 654
      S E + SP P
Sbjct: 642 SPAEVKSPATVKS P 655

```

Score = 253 (38.0 bits), Expect = 9.0e-18, P = 9.0e-18  
Identities = 115/364 (31%), Positives = 166/364 (45%)

Query:	110	EVTVPVVOQGSVAVKKVASAEIEPPSTEKPEPAKIOPPLVEETAATAKEAPRAE--ETHVOVQ-	167
		E PVV + A K + AE +P+P+ K PA+ P ++ A+A+ PAE ++ V+V+	
Sbjct:	705	EAKSPVPAKSPAEC--SPAECAPPAEAKSPAESAESPAEAKS--PAEAKSPVEVKS	762
Query:	168	PSTEETPDAAEAATAVAE--NSVKVQPPPAEEA--PL-VEFPAAEIQQPSAAE--SPSVELL	220
		P ++P E A A ++AE + K + P EE P V+ P + + P EE SP	
Sbjct:	763	PEKAKSPVKEGAKSLAEAKSPEKAKSPKVEIKPPAAVKSPEKAKSPMEKAKSPEKAKT	822
Query:	221	AEILPPSAEESPEEP--PAEILPPPAEKSPSVELLGEIRSPSAQKAPIE-VQPLPAE--	275
		+ P +A + EE PA+I P KSP+ E E +S+ ++ V P E	
Sbjct:	823	LDVKSPEAKTPEAKEARKPADIRSPQOVKSPAKE--EAKSPEKEETRTKVPAPKKEEVK	879
Query:	276	GALAEAPAKVEPPTVEETLAEVQPLLPPEAPREEAREQLSTAMETPAEEA-P-TEFQVQ-	333
		+EE AK P VEE E P P+ P +E+ ++ A + AEE P TE	
Sbjct:	880	SPVVEEAPKEKPPKVEE--EXTPATPKTVEKSSKKDEAPKEAQKPAKEKEPLTEKPKD	936
Query:	334	LPKETTAEESAIAEIQLLAATEPPADETPAE--ARSPLSEETSAAEEHA-EVQSPLAETT	390
		P E EEA + AA P +ETPA+ + + AE+A A E P +E	
Sbjct:	937	SPGEAKKEEAKEN--KAAA--PEEETPAKLGVKEEAKPKEAKDEADAKESPSPKSEKEP	991
Query:	391	A-EAESAIAEIQLLAIEAPADETPAEAQSPLESETSAEEAPAEVQSPSA-KGVSIEEAPLE	448
		EAE + E E+ + P + + EE Q PS K E++	
Sbjct:	992	KKEEVPAPEKDKTTEKTESKKPEKPKMQAKAKEEDKGLPQEPSKPKTEKAEKSST	105
Query:	449	LQPPSGEETTAEEESA 465	
		Q S A E A A	
Sbjct:	1052	DOKDSOPSEKAPEDKAA 1068	

Pedant information for DKFZphtes3 17f10, frame 3

Report for DKFZphtes3 17f10.3

```
[LENGTH]      710
[MW]           75131.94
[pI]           4.02
[KW]           All Alpha
[KW]           LOW COMPLEXITY      34.08 %
```

[illegible]



```
SEQ PAEARSPLSEETSAEEAAHVQSPLAEETTAEEASAIQLLAAIEAPADETPAEQAQSPIS
SEG xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx
PRD cccccccccchhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhccccc

SEQ EETSAEAPAEVQSPSAKGVSIIEAPLELQPSGTEETAEASAIQLLAAETAEAEAP
SEG xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx
PRD chhhhccccccccccccccccccccccccccccchhhhhhhhhhhhhhhhhhhhhhhhhhhc

SEQ AEVQPPPAEEAPAEVQPPPAEEAPAEVQPPPAEEAPAEVQPPPAEEAPAEVQPPPAEEAP
SEG xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx
PRD ccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc

SEQ AEVQPPPAEEAPAEVQPPPAEEAPAEVQSPPLAEETPIEETLAAVHSPADDVPAAEASVD
SEG xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx
PRD cccccccccccccccccccccccccccccccccccccccccchhhhhhhhhcccccccccccccc

SEQ KHSPPADLLITEEFPFGEASAEVSPPPSEQTPDEALVENVSTEFQSPQVAGIPAVKLGS
SEG xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx
PRD cccccccccccccccccccccccccccccccccccccccccchhhhhhhhhcccccccccccccc

SEQ VVLEGEAKFEVYSKINVLKDLSTNDQAPTLIESVFHLEIKQRPPEL
SEG eeehhhhhhhhccccceeeccccccccccccceehhhhhhhhhhhcccccc
PRD
```

(No Prosite data available for DKFZphtes3\_17f10.3)

(No Pfam data available for DKFZphtes3\_17f10.3)

DKFZphtes3\_17117

-----

group: metabolism

DKFZphtes3\_17117 encodes a novel 626 amino acid protein with similarity to transketolases (EC 2.2.1.1).

The novel protein contains a ATP/GTP-binding site motif A (P-loop). It is a new testis-specific transketolase. Transketolase requires thiamin pyrophosphate as cofactor and shows a wide specificity for both reactants, e.g. converts hydroxypyruvate and R-CHO into CO(2) and R-CHOH-CO-CH(2)OH.

The new protein can find application in modulation of metabolic pathways involving this transketolase activity and as a new enzyme for biotechnologic production processes.

strong similarity to transketolases

few EST hits (all from testis or pooled librarys containing testis)  
testis specific transketolase?

Sequenced by GBF

Locus: unknown

Insert length: 2688 bp

Poly A stretch at pos. 2649, polyadenylation signal at pos. 2630

```
1  GACAAAAGAG AGATGATGGC CAACGACGCC AAGCCCGACG TGAAGACCGT
51 GCAGGTGCTG CGGGACACAG CCAACCGCCT GCGGATCCAT TCCATCAGGG
101 CCACGTGTGC CTCTGGTTCT GGCCAGCTCA CGTCGTGCTG CAGTGCAGCG
151 GAGGTGCTGT CTGTCTCTCT TTCCACACG ATGAAGTATA AACAGACAGA
201 CCCAGAACAC CCGGACAAAG ACCGGTTCAT CCTCTCCAGG GGACATGCTG
251 CTCCTATCCT CTATGCTGCT TGGGTGGAGG TGGGTGACAT CAGTGAATCT
301 GACTTGCTGA ACCTGAGGAA ACTTCACAGC GACTTGAGAA GACACCCCTAC
351 CCCGCGATTG CCGTTTGTGG ACGTGGCAAC AGGGTCCCTA GGTGAGGGAT
401 TAGGTACTGC ATGTGGAATG GCTTATACCT GCAAGTACCT TGACAAGGCC
451 AGCTACCGGG TGTCTGCTCT TATGGGAGAT GGCGAATCCT CAGAAGGCTC
501 TGTGTGGGAG GCTTTTGCCT TTGCCCTCCA CTACAACCTG GACAATCTCG
551 TGGCGGTCTT CGACGTGAAC CGCTTGGGAC AAAGTGGCCC TGCACCCCTT
601 GAGCATGGCG CAGACATCTA CCAGAAATGC TGTGAAGCCT TTGGATGGAA
651 TACTTACTTA GTGGATGGCC ATGATGTGGA GGCCTTGTGC CAAGCATTTT
701 GGCAAGCAAG TCAAGTGAAG AACAAAGCCTA CTGCTATAGT TGCCAAGACC
751 TTCAAAGGTC GGGGTATTCC AAATATTGAG GATGCAGAAA ATTGGCATGG
801 AAAGCCAGTG CCAAAAGAAA GAGCAGATGC AATTGTCAAA TTAATTGAGA
851 GTCAGATACA GACCAATGAG AATCTCATAC CAAAATCGCC TGTGGAAGAC
901 TCACCTCAAA TAAGCATCAC AGATATAAAA ATGACCTCCC CACCTGCCTA
951 CAAAGTTGGT GACAAGATAG CTACTCAGAA AACATATGGT TTGGCTCTGG
1001 CTAAACTGGG CCGTGCAAAAT GAAAGAGTTA TTGTTCTGAG TGGTGACACG
1051 ATGAACCTCA CTTTTCTGTA GATATTCAGG AAAGAACACC CTGAGCGTTT
1101 CATAGAGTGT ATTATTGCTG AACAAAACAT GGTAAAGTGT GCACTAGGCT
1151 GTGCTACACG TGGTCGAACC ATTGCTTTTG CTGGTGCTTT TGCTGCCTTT
1201 TTTACTAGAG CATTGATGTA GCTCCGAATG GGAGCCATTT CTCAAGCCAA
1251 TATCAACCTT ATTGGTTCCC ACTGTGGGGT ATCCACTGGA GAAGATGGAG
1301 TCTCCAGATG GGCCTTGGAG GATCTAGCCA TGTTCGAAG CATTCCCAAT
1351 TGTACTGTTT TCTATCCAAG TGATGCCATC TCGACAGAGC ATGCTATTTA
1401 TCTAGCCGCC AATACCAAGG GAATGTGCTT CATTGCAACC AGCCAACCAAG
1451 AAAGTGCAGT TATTTATACC CCACAAGAAA ATTTTGAGAT TGGCCAGGCC
1501 AAGGTGGTCC GCCACGGTGT CAATGATAAA GTACAGTAA TTGGAGCTGG
1551 AGTTACTCTC CATGAAGCCT TAGAAGCTGC TGACCATCTT TCTCAACAAG
1601 GTATTTCTGT CCGTGTATC GACCAATTTA CCATTAAACC CTTGGATGCC
1651 GCCACCATCA TCTCCAGTGC AAAAGCCACA GCGGCGCAGG TTATCACAGT
1701 GGAGGATCAC TACAGGGAAG GTGGCATTGG AGAAGCTGTT TGTGCAGCTG
1751 TCTCCAGGGA GCCTGATATC CTTGTTTATC AACTGGCAGT GTCAGGAGTG
1801 CCTCAACGTG GGAATACTAG TGAATTGCTG GATATGTTTG GAATCAGTAC
1851 CAGACACATT ATAGCAGCCG TAACACTTAC TTTAATGAAG TAAACTAGGC
1901 TTATTTCTAA AAAGTCAAGT CTATTGGCTT TGGCCAAAAA GCACTGGTAT
1951 CTTTGATTTA AATTGATGTT TATTGTCACA AAACCATTTT TTATACCTAT
2001 ACAGTTGTAC TGTTCCTTTT AAAGCAAAGC CATTAAACAT CTTTCTTCAT
2051 TCCTAATTTG GAAATTAAG TTTACCTTTC TGTTAATCTA TGTATAAATG
2101 TTACTCTGAG TTATTAATGT GGATTTTAAA ATTGTAAGCA ATAGAATAGG
2151 AAATAAAACA ACTACCTAAT ACAAAATATT CTGATAAGAC TACAATATATC
2201 TGACTGAGCT GGGGATTTAA GTAGAGGTAA CTGTATCTTA AATGAGTATG
2251 ATTTCTTGT AAGTTAAAAA AATTGAAATT TAATTGTAGA CTTCAATAGT
2301 CCAAGTTTGT AAGGATGTTT GAGCTTTTGT ATAATGCCAT TTATACCTGC
2351 AGTTTTCAG ATAATGTTTG ACTGCAGTTG CCTTGGAAAT TCCTCCAAAG
2401 TTTGCCCTCA TCTCTCCTCT ACAGTTTGGG GGTGATGGTG CAGCAGTGGA
2451 ACATCTCTTG ATGCACCACA CTACTTGTGT TCTGTGAAGT GATGAAAGTA
```

2501 TAACTGGTTC TAGTTTGAC ACTACACACA TAGTTTGTG AAGCTTCAGA  
2551 AATGTTTTT CTTTCTCTG TGGCCAAACC AGTTTGTAA TCTGATTATA  
2601 TTCATCTGCT AATGATACTA AAGTTAATGT AATAAAGCAT TTAAAAATCA  
2651 GAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA

## BLAST Results

No BLAST result

## Medline entries

96214928:  
Amplification of the transketolase gene in desensitization-resistant  
mutant  
Y1 mouse adrenocortical tumor cells.

99123875:  
Properties and functions of the thiamin diphosphate dependent enzyme  
transketolase.

## Peptide information for frame 1

ORF from 13 bp to 1890 bp; peptide length: 626  
Category: strong similarity to known protein  
Classification: Metabolism  
Prosite motifs: ATP\_GTP\_A (595-603)

1 MMANDAKPDV KTVQVLRDTA NRLRIHSIRA TCASGSGQLT SCCSAAEVVS  
51 VLFHTMKYK QTDPEHPDND RFILSRGHAA PILYAAWVEV GOISESDLN  
101 LRKLHSDLER HPTPLPFVD VATGSLGQGL GTACGMAYTG KYLDKASYRV  
151 FCLMGDGESE EGSVWEAFV ASHYNLNLV AVFDVNLGQ SGPALEHGA  
201 DIYQNCCEAF GWNTYLVDPGH DVEALCQAFW QASQVKNKPT AIVAKTFKGR  
251 GIPNIEDAEN WHGKVPVKER ADAIVKLIES QIQTNENLIP KSPVEDSPQI  
301 SITDIKMTSP PAYKVVDKIA TQKTYGLALA KLGRANERVI VLSGDTMNST  
351 FSEIFRKEHP ERFIECIIAE QNMVSVALGC ATRGRTIAFA GAFAAFFTRA  
401 FDQLRMGAIS QANINLIGSH CGVSTGEDGV SQMALEDLAM FRIPNCTVF  
451 YPSDAISTEH AIYLAANTKG MCFIRTSQPE TAVIYTPQEN FEIGQAKVVR  
501 HGVNDKVTVI GAGVTLHEAL EAADHLSQGG ISVRVIDPFT IKPLDAATII  
551 SSAKATGGRV ITVEDHYREG GIGEAACAAV SREPDILVHQ LAVSGVPQRG  
601 KTELLEDMFG ISTRHITAAV TLTLMK

## BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKF2phtes3\_17l17, frame 1

SWISSPROT:TKT\_MOUSE TRANSKETOLASE (EC 2.2.1.1) (TK) (P68).., N = 1,  
Score = 2222, P = 2.5e-230

SWISSPROT:TKT\_RAT TRANSKETOLASE (EC 2.2.1.1) (TK).., N = 1, Score =  
2202, P = 3.3e-228

TREMBL:RN09256\_1 product: "transketolase"; Rattus norvegicus  
Sprague-Dawley transketolase mRNA, complete cds., N = 1, Score = 2202,  
P = 3.3e-228

SWISSPROT:TKT\_HUMAN TRANSKETOLASE (EC 2.2.1.1) (TK).., N = 1, Score =  
2200, P = 5.3e-228

>SWISSPROT:TKT\_MOUSE TRANSKETOLASE (EC 2.2.1.1) (TK) (P68).  
Length = 623

## HSPs:

Score = 2222 (333.4 bits), Expect = 2.5e-230, P = 2.5e-230  
Identities = 417/614 (67%), Positives = 501/614 (81%)

Query: 7 KPDVKTVQVLRDTANRLRIHSIRATCASGSGQLTSCCSAAEVVSVLFHTMKYKQTDPEH 66

Sbjct: KPD + +Q L+DTANRLRI SI+AT A+GSG TSCCSAAE+++VLFFHTM+YK DP +  
 6 KPDQQLQALKDNTANRLRISSIQATTAAGSGHPTSCCSAAEIMAVLFFHTMRYKALDPRN 65

Query: 67 PDNDRFILSRGHAAPILYAAMVEVGDISESDLLNLRKLHSDLERHPTPLPFVDVATGSL 126  
 P NDRF+LS+GHAAPILYA W E G + E++LLNLRK+ SDL+ HP P+ F DVATGSL

Sbjct: 66 PHNDRFVLSKGHAAPILYAVWAEAGFLPEAELLNLRKISSDLDGHPVPKQAFDVTATGSL 125

Query: 127 GQGLGTACGMAYTGKYLKASYRVFCLMGDGESEGSVWEAFASFASHYNLDNLVAVFDVN 186  
 M +PP+YKVGDKIAT+K YGLALAKL A++R+I L GDT NSTFSE+F+KEHP+RFIEC

Sbjct: 126 GQGLGAACGMAYTGKYLKASYRVFCLMGDGESEGSVWEAMAFAGIYKLDNLVAIFDIN 185

Query: 187 RLQSGGPAPLEHGADIYQNCCEAFGWNTYLVGDHDEALCQAFWQASQVKNKPTAIVAKT 246  
 RLQGS PAPL+H DIYO CEAFGW+T +VDGH VE LC+AF QA K++PTAI+AKT

Sbjct: 186 RLQGS DPAPLQHQVDIYQKRCEAFGWHTIIVDGHSEELCKAFQA---KHQPTAIIAKT 242

Query: 247 FKGRGIPNIEDAENWHGKVPKERADAIVKLIESIQITNENLIPKSPVEDSPQISITDIK 306  
 FKGRGI IED E WHGK+PK A+ I++ I SQ+Q+ + ++ P ED+P + I +I+

Sbjct: 243 FKGRGITGIEDKEAWHGKPLPKNMAEQIIQEIYSQVSKKILATPPQEDAPSVDIANIR 302

Query: 307 MTPSPAYKVGDKIATQKTYGLALAKLGRANERVIVSGDTMNSTFSEIFRKEHPERFIEC 366  
 M +PP+YKVGDKIAT+K YGLALAKL A++R+I L GDT NSTFSE+F+KEHP+RFIEC

Sbjct: 303 MTPSPSYKVGDKIATRKAYGLALAKLGHASDRIIALDGTNKNSTFSEIFRKEHPERFIEC 362

Query: 367 IIAEQNMVSVAGCATRGRTIAFAGAFAAFFTRAFDQLRMGAISQANINLIGSHCGVSTG 426  
 IAEQNMVS+A+GCATR RT+ F FAAFFTRAFDQ+RM AIS++NINL GSHCGVS G

Sbjct: 363 YIAEQNMVSVAGCATRDRTVFFCSTFAAFFTRAFDQIRMAAISQANINLIGSHCGVSTG 422

Query: 427 EDGVSQMALEDLAFMRSPNCTVFYPSDAISTEHAIIYLAANTKGMCFIRTSQPETAIVIYT 486  
 EDGVSQMALEDLAFMRSP TVFYPSD ++TE A+ LAANTKG+CFIRTS+PE A+IY+

Sbjct: 423 EDGVSQMALEDLAFMRSPVPMSTVFYPSDGVATEKAVELAANTKGICFIRTSRPNIAIYS 482

Query: 487 PQENFEIGQAKVVRHGVNDKVTIVIGAGVTLHEALEADHLSQGGISVRVIDPFTIKPLDA 546  
 E+F++QAKVV +D+VTIVIGAGVTLHEAL AA+ L + IS+RV+DPFTIKPLD

Sbjct: 483 NNEDFQVQAKVVLKSKDDQVTIVIGAGVTLHEALAAEESLKKDKISIRVLDPFTIKPLDR 542

Query: 547 ATIISAKATGGRVITVEDHYREGGIGEAACAASREPDIIVHQLVAGSVGPQRGKTSSELL 606  
 I+ SA+AT GR++TVEDHY EGGIGEA AAV EP + V +LAVS VP+ GK +ELL

Sbjct: 543 KLILDSARATKGRILTVEDHYEGGIGEAASAAVGEVGPVTVTRLAVSQVPRSGKPAELL 602

Query: 607 DMFGISTRHIIAAV 620  
 MFGI I+ AV

Sbjct: 603 KMFGIDKDAIVQAV 616

Pedant information for DKFZphtes3\_17117, frame 1

Report for DKFZphtes3\_17117.1

[LENGTH] 626  
 [MW] 67877.52  
 [PI] 5.90  
 [HOMOL] SWISSPROT:TKT\_MOUSE TRANSKETOLASE (EC 2.2.1.1) (TK) (P68). 0.0  
 [FUNCAT] m outer membrane and cell wall [M. jannaschii, MJO681] 3e-48  
 [FUNCAT] g carbohydrate metabolism and transport [H. influenzae, HI1023] 9e-36  
 [FUNCAT] 01.05.01 carbohydrate utilization [S. cerevisiae, YPR074c] 5e-32  
 [FUNCAT] 30.03 organization of cytoplasm [S. cerevisiae, YPR074c] 5e-32  
 [FUNCAT] 02.07 pentose-phosphate pathway [S. cerevisiae, YPR074c] 5e-32  
 [FUNCAT] 01.01.01 amino-acid biosynthesis [S. cerevisiae, YPR074c] 5e-32  
 [FUNCAT] i lipid metabolism [H. influenzae, HI1439] 3e-17  
 [FUNCAT] c energy conversion [H. influenzae, HI1233] 2e-09  
 [FUNCAT] 02.01 glycolysis [S. cerevisiae, YBR221c PDB1 - pyruvate dehydrogenase] 2e-05  
 [FUNCAT] 30.16 mitochondrial organization [S. cerevisiae, YBR221c PDB1 - pyruvate dehydrogenase] 2e-05  
 [BLOCKS] BL00801F  
 [BLOCKS] BL00801E  
 [BLOCKS] BL00801D Transketolase proteins  
 [BLOCKS] BL00801C Transketolase proteins  
 [BLOCKS] BL00801B Transketolase proteins  
 [BLOCKS] BL00801A Transketolase proteins  
 [SCOP] dtrka2 3.28.1.2.1 Transketolase Transketolase, C-terminal domai 1e-21  
 [EC] 1.2.4.1 Pyruvate dehydrogenase (lipoamide) 8e-11  
 [EC] 1.2.4.4 3-Methyl-2-oxobutanoate dehydrogenase (lipoamide) 4e-10  
 [EC] 2.2.1.1 Transketolase 0.0  
 [EC] 2.2.1.3 Formaldehyde transketolase 1e-20  
 [PIRKW] transferase 0.0  
 [PIRKW] flavoprotein 2e-07  
 [PIRKW] Calvin cycle 1e-40  
 [PIRKW] heterotetramer 2e-07

{PIRKW} pentose phosphate pathway 0.0  
 {PIRKW} magnesium 1e-40  
 {PIRKW} thiamine pyrophosphate 0.0  
 {PIRKW} oxidoreductase 7e-12  
 {PIRKW} fatty acid biosynthesis 4e-10  
 {PIRKW} mitochondrion 2e-07  
 {PIRKW} peroxisome 1e-20  
 {PIRKW} homodimer 1e-40  
 {SUPFAM} pyruvate dehydrogenase (lipoamide) alpha chain 1e-06  
 {SUPFAM} pyruvate dehydrogenase (lipoamide) beta chain 7e-12  
 {SUPFAM} ferredoxin 2[4Fe-4S]-related protein 8e-47  
 {SUPFAM} thiamine pyrophosphate-binding domain homology 0.0  
 {SUPFAM} pyruvate dehydrogenase (lipoamide) 6e-08  
 {SUPFAM} ferredoxin 2[4Fe-4S] homology 8e-47  
 {SUPFAM} hypothetical protein C2814 2e-21  
 {SUPFAM} transketolase 0.0  
 {PROSITE} ATP\_GTP\_A 1  
 {PFAM} Transketolase  
 {KW} Alpha\_Beta  
 {KW} 3D  
 {KW} LOW\_COMPLEXITY 3.04 %

SEQ MMANDAKPDVKTQVLRDTANRLRIHSIRATCAGSGQLTSCCSAAEVSVLFHTMKYK  
 SEG .....  
 lngsB .....HHHHHHHHHHHCCCHHHHHHHHHHHHHHHHHH-HHCCCT  
  
 SEQ QTDPEHPDNDRFILSRGHAAPILYAAWVEVGDISDLLNLRLKLSDLERHPTPRLPFVD  
 SEG .....  
 lngsB TTTTTTTTCEEEETGGGHHHHHHHHHHCTTCHHHHTTTTTTTTTTTTTTTTTTTTC  
  
 SEQ VATGSLGQGLGTACGMAYTGKYLKASYRVFCLMGDSSEGSWEAFASFASHYNLDNLV  
 SEG .....  
 lngsB CCCCCCTHHHHHHHHHHHHHHHCBTTBTTEEECHHHHHCHHHHHHHHHHHHCTTTEE  
  
 SEQ AVFDVNRLLGQSGPAPLEHGADIYQNCCEAFGWNTYLVGDHVEALCQAFWQASQVKNKPT  
 SEG .....  
 lngsB EEEEECCETTEEEGGCCCCCHHHHH-HHHCCEEEETTTTTHHHHHHHHHHHHTTTTCE  
  
 SEQ AIVAKTFKGRGIPNIEDAENWHGKVPKERADAIVKLIESQIQTNENLIPKSPVEDSPQI  
 SEG .....  
 lngsB EEEEECTTTTTTCHHHHHHHHHHTCCHHHHHHHHHHHHHHHHHHHHHHHHHHHHHCHHH  
  
 SEQ SITDIKMTSPPAYKVGDKIATQKTYGLAKLGRANERVIVLSGDTMNSTFSEIFRKEHP  
 SEG .....  
 lngsB HHHHHHHHTCCCTTTTCBCHHHHHHHHHHHHTTTTTTEEEETTHHHHCCTTCEEECCG  
  
 SEQ ERFIECIIAEQNMVSVLGCATRGRTIAFAGAFAAFTRAFQDLRMGAISQANINLIGSH  
 SEG .....  
 lngsB GCEETTTTHHHHHHHHHHHHTTTTTTEEEEGGGGGGHHHHHHHHHHCTTTEEEEC  
  
 SEQ CGVSTGEDGVSMALDLAMFRSIPNCTVFYPSDAISTEHAITYLAANTKGMCFIRTSQPE  
 SEG .....  
 lngsB CCGGGTTTTTTTCHHHHHHHCTTTTEEECCCHHHHHHHHHHTTTTCEEECCCCCB  
  
 SEQ TAVIYTPQENFEIGQAKVVRHGVNDKVTIVIGAGVTLHEALEADHLSQOGISVRVIDPFT  
 SEG .....  
 lngsB CCTTTTCHHHHHCC-CEEEETTTTTTEEEECCHHHHHHHHHHHHHHHHCCCEEE...  
  
 SEQ IKPLDAATIISSAKATGGRVITVEDHYREGGIGEAVCAVSREPDILVHQLAVSGVPQRG  
 SEG .....  
 lngsB .....  
  
 SEQ KTSELDMFGISTRHIIAAVTLTLMK  
 SEG .....  
 lngsB .....

Prosite for DKF2phtes3\_17117.1

PS00017 595->603 ATP\_GTP\_A PDOC00017

Pfam for DKF2phtes3\_17117.1

HMM\_NAME Transketolase  
 HMM \*vNtIriLaMDAVEKANSQHPGaPMGNAPMAHVLWqrMMRHNPNDPrWPN

Query	20	+N++RI ++ A + +SG +++++A++ VL++++M++++DP P+	68
HMM		ANRLRIHSIRATCASGSGQLTSCCSAAEVSVLFFHTMKYKQTDPEHPD	
Query	69	RDRFVLSNGHaCMLLYsMWHLyGYDMPmWDLKQFRQWHSrTPGHFEIghT	117
HMM		+DRF+LS GH++++LY+ W + G +++++DL++++R++HS++ +HP ++	
Query	118	NDRFILSRGHAAPILYAAMVEVD-ISESDLNLRKLHSDLERHPTPLRP	157
HMM		PGVEVTTGPIGGGIANaVWMAIAERNLAATYNRPFGDIfDHYTYCFMGDG	
Query	158	++ +V+TG+LGQG++ +++++Y++++ D++++++C+MGDG	207
HMM		FV-DVATGSLGGGLG-----TACGMAYTKYLDKASYRVFCLMGDG	
Query	208	CLMEGISWEACSLAGHMqLGNWIAFYDDNrISIDGdTDfWFqEDtYakRF	255
HMM		+ +EG++WEA ++A+H++L+N++A +D NR++++G++++ + D+Y+ +	
Query	256	ESSEGSVWEAFAFASHYNLDNLVAVFDVNRLGQSGPAPLEHGADIYQNC	269
HMM		EAYGWHVIEVENDGHdVBeIcaAIEeAKaekDRPTLIICRTVIGYGSPlk	
Query	311	EA+GW++ +V DGHdVE++C A+ +A +K++PT+I ++T++G+G+PN	358
HMM		EAFGWNTYLV--DGHdVEALCQAFQASQVKNKPTAIVAKTFKGRGIPNI	
Query	359	QGTdHWHGAPLGeD*	392
HMM		++ + WHG+P +++	
Query	443	EDAENWHGKVPVKE	490
HMM		*PqWePnddkIATRKASQqLeaiGPaLPefWGGSADLTSPNLTrWKGmv	
Query	491	P++++ +DKIAT K+++ AL+++G A +++ +S+D+ +S+++++ ++	538
HMM		PAYKV-GDKIATQKTYGLALAKLGRANERVILSGDTMNSTFSEIFRKE	
Query	539	WFMPPSISTDCynGNwsGRYIHGYIREHgmAIMNGIALHGnFRPYGGT	587
HMM		+ + R+I++ I+E++M++++ G+A++G+ +++++ G	
Query	587	H-----PERFIECIIAEQNMVSVALGCATRGR-TIAFAGA	616
HMM		FMMFyDYARPAIRMAALMeLPVIVWVTHDSIGLGEDGPTHQPVEHLAHR	
Query	616	F++F+++A++++RM A++ ++++++H++++ GEDG +++++E+LA+FR	
HMM		FAAFFTRAFQQLRMGAISQANINLIGSHCGVSTGEDGVSMQALDAMFR	
Query	642	aIPNMsvWRPCDgNETayAWyIAvErehTPtILILSRQNLPLQIErNPrqf	
HMM		+IPN +V++P+D+ T+ A YLA+++++ +++++S ++ +++++ P +	
Query	642	SIPNCTVFYPSDAISTEHAIIAANTKGM-CFIRTSQPETAIVIY-PQEN	
HMM		ekvaRGGYVLkOmdnePDVILIAIGSEMELavaAAKLladEGIKaRVVSM	
Query	642	+++++++V + + + + V+I++G++++A++AA+ L+ +GI +RV+++	
HMM		FEIGQAKVVRHGVN--DKVTVIGAGVTLHEALEAADHLSQGGISVRVIDP	
Query	642	PCTeWFD.....kQDeEYReSVLPdhVPqRVaVEmGvtWCWYKYVGqq	
HMM		++++++D +++++R +DH++ +++++V ++ ++ +	
Query	642	FTIKPLDAATIISSAKATGGRVITVEDHYR-EGGIGEAVCAAVSREPIL	
HMM		GaIfGMN+FGESSGKAPpevLYkMFgFTPENI*	
Query	642	+ +++ ++ + + + +L+ MFG+ +I	
HMM		VHQLAVSGVPQR---GKTSELDMFGISTRHI	

DKFZphtes3\_17n12  
-----

group: transcription factors

DKFZphtes3\_17n12.1 encodes a novel 804 amino acid protein which is nearly identical to mouse and trout SOX-LZ.

Sox proteins belong to the HMG box superfamily of DNA-binding proteins and are involved in the regulation of developmental processes as germ layer formation, organ development and cell type specification. Deletion or mutation of Sox proteins often results in developmental defects and congenital disease in humans. Sox proteins perform their function in a complex interplay with other transcription factors in a manner highly dependent on cell type and promoter context. The new protein is related to the SOX-LZ protein and contains an additional leucine-zipper.

The new protein can find application in modulating/blocking the expression of SOX-controlled genes.

nearly identical to mouse SOX-LZ

complete cDNA, complete cds, few EST hits  
mouse and trout SOX-LZ, involved in spermatogenesis

Sequenced by GBF

Locus: unknown

Insert length: 2802 bp

Poly A stretch at pos. 2692, polyadenylation signal at pos. 2660

```

1 GGGATAGGAA AGATGAAAGG TCATGGTGAG CTTCAAGGAC ATGAAAGGTT
51 GTTGCTCAT GTAAACAATAG TAGATTGTTT TTTTCTCTAA TATTTCTAGC
101 CAGCCCTTAA GTCAGGTGAT GGAACAATA CCTACAGTTT AGTCAGGTGA
151 AACAGGAGTG GGTGGAGGAA GGAAGAAGA AAAATGGGAA GAATGCTTTC
201 CAAGCAAGCC ACCTCTCCAT TTGCTGTGTC AGCTGATGGA GAGGATGCAA
251 TGACCCAGGA TTTAACCTCA AGGGAAGAGG AAGAGGGCAG TGATCAACAT
301 GTGGCTCCCT ATCTGCCTCT GCACCCCATTA ATGCACAACA AACCTCACTC
351 TGAGGAGCTA CCAACACTTG TCAGTACCAT TCAACAGAGT GCTGACTGGG
401 ACAGCGTTCT GTCACTCTCG CAAGAATGAG AATCAGAGAT TAATACTTTA
451 TGTTCCTTAT ATTCTCTCCG AATATCCTCT ACCTCACCAC ATAGCCTGA
501 CGAAGGGAGT CGGGACCGTG AGATAATGAC CAGTGTACTT TTTGAAACCC
551 CAGAGCGGCG CAAGGGGACT CTTGCCGATG TGTGGACAC ACTGAAACAG
601 AGAAGCTTGT AGGAATGAC TCGGACTGAA CAAGAGGATT CTCTCTCAT
651 GGAATACTA CTTTCAAAAG ATTGCAAGGA AAAAATGGAA AGACTAAATA
701 CCAGTCAACT TCTTGGACAA ATTAAAGGTA CACCTGAGAG CCTGGCAGAA
751 AAAGAACGGC AGCTCTCCAC CATGATTACC CAGCTGATCA GTTTACGGGA
801 CCAGCTACTG GCAGCGCATG ATGAACAGAA AAAACTGGCA CGCTCACAAA
851 TTCAGAAACA ACGGACGCAA ATGGACCTTG CTCGCCACA GCAAGAACAG
901 ATTGCGAGAC AACAGCAGCA ACTTCTGCAA CAGCAGCACA AAATTAATCT
951 CCTGCAGCAA CAGATCCAGG TTCAGGGTCA CATGCTCCG CTCTATGATC
1001 CAATTTTCC ACATGACCAG CGGACTCTGG CAGCAGCTGC TGCTGCCCAA
1051 CAGGGATTCC TCTTCCCCC TGGAAATAA TACAACACAG GTGATACTA
1101 CCCCCTACAG TTCAATTCCAT CAACAATGGC AGCTGCTGCT GCTTCTGGAC
1151 TCAGCCCTTT ACAGCTCCAG CAGCTCTATG CCGCTCAGCT GGCCAGCATG
1201 CAGGTGTACG CTGGAGCAAA GATGCCATCA ACTCCACAGC CACCAAACAC
1251 AGCAGGGACG GTCTCACCTA CTGGGATAAA AAATGAAAG AGAGGGACCA
1301 GCCCTGTAAC TCAAGTTAAG GATGAAGCAG CAGCAGACCC TCTGAATCTC
1351 TCATCCCGAC CCAAGCAGC AGAGCCTGTA AAGTCCCAA CGTCTCCAC
1401 CCAGAACCTC TCCCAGCCA GCAAAACCAG CCCTGTCAAT CTGCCAAA
1451 AAAGCAGCAT CCCTAGCCCC ATTGGAGGAA GCCTGGGAAG AGGATCCTCT
1501 TTAGGTAAAT GGAAGTCA ACACCAGGAA GAGACTTACG AATTAGATAT
1551 CTAATCTAGT CTCAACTCCC CTGCCCTTTT TGGGGATCAG GATACAGTGA
1601 TGAAAGCCAT TCAGGAGGCG CGGAAGATGC GAGAGCAGAT CCAGCGGGAG
1651 CAACAGCAGC AACAGCCACA TGGTGTGAC GGGAAACTGT CCTCCATAAA
1701 TAATATGGGG CTGAACAGCT GCAGGAATGA AAAGGAAAGA ACGGCTTTG
1751 AGAATTTGGG GCCCCAGTTA ACGGGAAAGT CAAATGAAGA TGGAAACTG
1801 GGCCAGGTG TCATCGACCT TACTCGGCCA GAAGATGCAG AGGGAAGTAA
1851 AGCAATGAAT GGCTCTGCG CTAACATACA GCAGTATTAT TGTGGCCAA
1901 CAGGAGGTGC CACTGTGGCT GAAGCACGAG TCTACAGGGA CGCCCGCGGC
1951 CGTGCCAGCA GCGAGCCACA CATTAGCGA CCAATGAATG CATTATGGT
2001 TTGGGCAAG GATGAGAGGA GAAAAATCCT TCAGGCCTTC CCCGACATGC
2051 ATAACCTCAA CATTAGCAA ATCTTAGGAT CTGCTGGAA ATCAATGTCC
2101 AACCCAGGAG AGCAACCTTA TTATGAAGAG CAGGCCCGGC TAAGCAAGAT
2151 CCACTTAGAG AAGTACCCAA ACTATAAATA CAAACCCGCA CCGAAACGCA
2201 CTGCAATTGT TGATGGCAA AAGCTTCGGA TTGGGGAGTA TAAGCAACTG
2251 ATGAGGTCTC GGAGACAGGA GATGAGGCG TTCTTTACTG TGGGGCAAC
2301 GCCTCAGATT CCAATCACA CAGGAACAGG TGTGTGTAT CTGGTGCTA
2351 TCACTATGGC AACTACCACA CCATCGCCTC AGATGACATC TGACTGCTCT

```

```

2401 AGCACCTCGG CCAGCCCGGA GCCCAGCCTC CCGGTCATCC AGAGCACTTA
2451 TGGTATGAAG ACAGATGGCG GAAGCCTAGC TGGAAATGAA ATGATCAATG
2501 GAGAGGATGA AATGGAAATG TATGATGACT ATGAAGATGA CCCCAAATCA
2551 GACTATAGCA GTGAAATGA AGCCCGGAG GCTGTCAGTG CCAACTGAGG
2601 AGTTTTTGTG TGCTGAATTA AAGTACTCTG ACATTTCACC CCCCTCCCA
2651 ACAAAGAGTT ATTAAGAGC CCGCATGCAT TTGTGGCTCC ACAATTAAAA
2701 AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA
2751 AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA
2801 AA

```

## BLAST Results

No BLAST result

## Medline entries

95311974:  
A gene that is related to SRY and is expressed in the testes encodes a leucine zipper-containing protein.

96032826:  
The Sry-related HMG box-containing gene Sox6 is expressed in the adult testis and developing nervous system of the mouse.

## Peptide information for frame 1

ORF from 184 bp to 2595 bp; peptide length: 804  
Category: strong similarity to known protein

```

1 MGRMSSKQAT SPFACAADGE DANTQDLTSR EKEEGSDQHV ASHLPLHPIM
51 HNKPHSEELP TLVSTIQQDA DWDSVLSSQQ RMESENKLC SLYSFRNTST
101 SPHKPOEGSR DREIMTSVTF GTPERRKGS LADVVDLAKK KLEEMTRTEQ
151 EDSSCEKELL SKDWKEKMER LNTSELLGEI KGTPESLAEK ERQLSTMHTQ
201 LISLREQLLA AHDEQKLLAA SQIEKQRQOM DLARQQEQEI ARQQOQLLQ
251 QHKINLLQQQ IQVQGHMPPL MIPFFPHDQR TLAAAAAQQ GLFFPFCITY
301 KPGDNYPVQF IPSTMAAAA SGLSPLQLQQ LYAAQLASHQ VSPGAKMPST
351 PQPPNTAGTV SPTGIKNEKR GTSFVTVKDD EAAAOPLNLS SRPKTAEPVK
401 SPTSPTQNLF PASKTSFQND TWKAIQEAR KMREQIQREQ QQQPHGVGD
451 TYELDILSSL NSPALFGDQD TWKAIQEAR KMREQIQREQ QQQPHGVGD
501 KLSSINNMGL NSCRNEKERT RFENLGPQLT GKSNEQCKLG PGVIDLTRPE
551 DAEGSKAMNG SAAKLQYYC WPTGGATVAE ARVYRDARGR ASSEPHIKRP
601 MNAFNIHAKD ERRKILQAF DMHNSNISKI LGRWKSMSN QEQQYVEEQ
651 ARLSKIHLEK YPNYKYKPRP KRTCIVDCKK LRIGEYKQIM RSRQEMRQF
701 FTVQQQPQIP ITTGTGVVYP CAITMATTTP SPQMTSDCSS TSASPEPSLP
751 VIQSTYGMKT DGGSLAGNEM INGEDEMEMY DDYEDDPKSD YSSENEAPEA
801 VSAH

```

## BLASTP hits

Entry MMSOXL22.1 from database TREMBL:  
product: "SOX-L2"; Mouse mRNA for SOX-L2, complete cds.  
Score = 3910, P = 0.0e+00, identities = 764/801, positives = 774/801

Entry I51083 from database PIR:  
SOX-L2 - rainbow trout  
Score = 1774, P = 1.1e-287, identities = 365/532, positives = 431/532

Entry S59121 from database PIR:  
SOX6 protein - mouse  
Score = 2319, P = 1.2e-240, identities = 489/660, positives = 527/660

Entry AB006330.1 from database TREMBL:  
gene: "mSox5L"; product: "SOX5"; Mus musculus mSox5L mRNA, complete cds.  
Score = 1212, P = 8.9e-209, identities = 274/457, positives = 324/457

Entry MMU010604.1 from database TREMBL:  
gene: "sox5"; product: "L-Sox5 protein"; Mus musculus mRNA for transcription factor L-Sox5  
Score = 879, P = 4.2e-195, identities = 190/281, positives = 218/281



Alert BLASTP hits for DKFZphtes3\_17n12, frame 1  
No Alert BLASTP hits found

Pedant information for DKFZphtes3\_17n12, frame 1

Report for DKFZphtes3\_17n12.1

[LENGTH] 804  
[MW] 89332.69  
[pI] 6.97  
[HOMOL] TREMBL:MMSOXLZ2\_1 product: "SOX-LZ"; Mouse mRNA for SOX-LZ, complete cds. 0.0  
[FUNCAT] 04.05.01.04 transcriptional control [S. cerevisiae, YKL032c] 8e-07  
[FUNCAT] 30.10 nuclear organization [S. cerevisiae, YKL032c] 8e-07  
[FUNCAT] 01.07.07 regulation of vitamins, cofactors, and prosthetic groups [S. cerevisiae, YPR065w] 5e-06  
[FUNCAT] 03.04 budding, cell polarity and filament formation [S. cerevisiae, YBR089c-a] 7e-06  
[FUNCAT] 30.13 organization of chromosome structure [S. cerevisiae, YBR089c-a] 7e-06  
[FUNCAT] 03.01 cell growth [S. cerevisiae, YBR089c-a] 7e-06  
[FUNCAT] 03.16 dna synthesis and replication [S. cerevisiae, YMR072w] 2e-04  
[FUNCAT] 30.16 mitochondrial organization [S. cerevisiae, YMR072w] 2e-04  
[SCOP] d1hmf\_1.20.1.1.1 HMG1, fragments A and B [rat/hamster (Rattus)] 1e-13  
[SCOP] d1lefa\_1.20.1.1.6 Lymphoid enhancer-binding factor, LEF1 [mouse] 4e-15  
[SCOP] d1hrya\_1.20.1.1.4 SRY [Human (Homo sapiens)] 7e-17  
[PIRKE] DNA binding 4e-94  
[PIRKE] T-cell receptor 4e-07  
[PIRKE] leucine zipper 1e-38  
[PIRKE] alternative splicing 2e-07  
[PIRKE] transcription factor 4e-16  
[PIRKE] transcription regulation 1e-12  
[SUPFAM] HMG box homology 0.0  
[SUPFAM] unassigned HMG box proteins 4e-94  
[PROSITE] ATP\_GTP\_A 1  
[PROSITE] LEUCINE\_ZIPPER 1  
[PROSITE] MYRISTYL 6  
[PROSITE] AMIDATION 1  
[PROSITE] CAMP\_PHOSPHO\_SITE 2  
[PROSITE] CK2\_PHOSPHO\_SITE 14  
[PROSITE] PKC\_PHOSPHO\_SITE 10  
[PROSITE] ASN\_GLYCOSYLATION 6  
[PFAM] HMG (high mobility group) box  
[KW] Irregular  
[KW] 3D  
[KW] LOW\_COMPLEXITY 13.81 %  
[KW] COILED\_COIL 3.48 %

SEQ MGRMSSQKATSPFACAADGEDAMTQDLTSREKEEGSDQHVASHLPLHPIMHNKPHSEELP  
SEG .....  
COILS .....  
lnhm- .....  
SEQ TLVSTIQDADWDVLSQQRMESNNKLCSLYSFRNTSTSPHKPDEGSRDREIMTSVTF  
SEG .....  
COILS .....  
lnhm- .....  
SEQ GTPERRKGLADVVDTLKQKKLEEMTRTEQEDSSCKEKLKSKDWKEKMERLNTSELLGEI  
SEG .....  
COILS .....  
lnhm- .....  
SEQ KGTPESLAEKERQLSTMITQLISLREQLLAHDEQKKLAASQIEKQRQOMDLARQQQEQI  
SEG .....  
COILS .....  
lnhm- .....  
SEQ ARQQQQLLQQQHKNLQQQIQVQGHMPLMIPFPHDQRTLAAAAAQQGFLFPPPGITY  
SEG .....  
COILS .....  
lnhm- .....  
SEQ KPGDNYPVQFIPSTMAAAASGLSPQLQLYAAQLASMQVSPGAKMPSTPQPPNTAGTV  
SEG .....  
COILS .....  
lnhm- .....

```

COILS .....
lnhm- .....

SEQ      SPTGIKNEKRGTSPVTQVKDEAAQPLNLSSRPKTAEPVKSPTSPTQNLFPASKTSPVNL
SEG      .....
COILS    .....
lnhm-    .....

SEQ      PKNSSIPIGSSLGRGSSLGKWKSOHQEETVELDILSSLNSPALFGDQDTVMKAIQEAR
SEG      ...XXXXXXXXXXXXXXXXXXXXX.....
COILS    .....
lnhm-    .....

SEQ      KMREQIQREQQQQPHGVDGKLSSINNMGLNSCRNEKERTRFENLGPQLTGKSNEDGKLG
SEG      ..XXXXXXXXXXXXXXXXX.....
COILS    .....
lnhm-    .....

SEQ      PGVIDLTRPEDAEGSKAMNGSAAKLQOYYCWPTGGATVAEARVYRDARGRASSEPHIKRP
SEG      .....
COILS    .....
lnhm-    .....CCC

SEQ      MNAFMVWAKDERRKILQAFPMHNSNISKILGSRWKSMSNQEQPYEEQARLSKIHLEK
SEG      .....
COILS    .....
lnhm-    CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC

SEQ      YPNYKYKPRPKRTCIVDGKKLRIGEYKQLMRSRRQEMRQFFTVGQQPQIPITTTGTGVVYP
SEG      XXXXXXXXXXXXX.....
COILS    .....
lnhm-    HHHTTTTTTTT.....

SEQ      GAITMATTTTSPQMTSDCSSTSASPEPSLPVIQSTYGMKTDGGSLAGNEMINGEDEMEMY
SEG      .....
COILS    .....
lnhm-    .....

SEQ      DDYEDDPKSDYSSNEAPEAVSAN
SEG      XXXXXXXX.....
COILS    .....
lnhm-    .....

```

## Prosite for DKFZphtes3\_17n12.1

PS00001	97->101	ASN_GLYCOSYLATION	PDOC00001
PS00001	172->176	ASN_GLYCOSYLATION	PDOC00001
PS00001	388->392	ASN_GLYCOSYLATION	PDOC00001
PS00001	422->426	ASN_GLYCOSYLATION	PDOC00001
PS00001	559->563	ASN_GLYCOSYLATION	PDOC00001
PS00001	626->630	ASN_GLYCOSYLATION	PDOC00001
PS00004	126->130	CAMP_PHOSPHO_SITE	PDOC00004
PS00004	369->373	CAMP_PHOSPHO_SITE	PDOC00004
PS00005	5->8	PKC_PHOSPHO_SITE	PDOC00005
PS00005	28->31	PKC_PHOSPHO_SITE	PDOC00005
PS00005	94->97	PKC_PHOSPHO_SITE	PDOC00005
PS00005	136->139	PKC_PHOSPHO_SITE	PDOC00005
PS00005	203->206	PKC_PHOSPHO_SITE	PDOC00005
PS00005	299->302	PKC_PHOSPHO_SITE	PDOC00005
PS00005	390->393	PKC_PHOSPHO_SITE	PDOC00005
PS00005	512->515	PKC_PHOSPHO_SITE	PDOC00005
PS00005	530->533	PKC_PHOSPHO_SITE	PDOC00005
PS00005	692->695	PKC_PHOSPHO_SITE	PDOC00005
PS00006	28->32	CK2_PHOSPHO_SITE	PDOC00006
PS00006	129->133	CK2_PHOSPHO_SITE	PDOC00006
PS00006	146->150	CK2_PHOSPHO_SITE	PDOC00006
PS00006	148->152	CK2_PHOSPHO_SITE	PDOC00006
PS00006	154->158	CK2_PHOSPHO_SITE	PDOC00006
PS00006	186->190	CK2_PHOSPHO_SITE	PDOC00006
PS00006	203->207	CK2_PHOSPHO_SITE	PDOC00006
PS00006	221->225	CK2_PHOSPHO_SITE	PDOC00006
PS00006	520->524	CK2_PHOSPHO_SITE	PDOC00006
PS00006	533->537	CK2_PHOSPHO_SITE	PDOC00006
PS00006	547->551	CK2_PHOSPHO_SITE	PDOC00006
PS00006	577->581	CK2_PHOSPHO_SITE	PDOC00006
PS00006	639->643	CK2_PHOSPHO_SITE	PDOC00006
PS00006	793->797	CK2_PHOSPHO_SITE	PDOC00006
PS00008	182->188	MYRISTYL	PDOC00008
PS00008	431->437	MYRISTYL	PDOC00008

PS00008	437->443	MYRISTYL	PDOC00008
PS00008	509->515	MYRISTYL	PDOC00008
PS00008	575->581	MYRISTYL	PDOC00008
PS00008	762->768	MYRISTYL	PDOC00008
PS00009	677->681	AMIDATION	PDOC00009
PS00017	526->534	ATP_GTP_A	PDOC00017
PS00029	187->209	LEUCINE_ZIPPER	PDOC00029

Pfam for DKFZphtes3\_17n12.1

HMM_NAME	HMG (high mobility group) box		
HMM	*PKRPMNAYMLWMQEMReIKaENPNdMhNtEISKMIgEMWKnMsEEEkM +KRPMNA+M+W+++ R+KI + P DMHN++ISK++G +WK+MS +EK+		
Query	597	IKRPMNAFMVWAKDERRKILQAFP-DMHNSNISILGSRWKSMSNQEKQ	644
HMM	PYEdMAeeEKqRYMKEMPeYK*		
Query	645	PYYEEQARLSKIHLEKYPNYK	665

DKF2phtes3\_17n18

group: intracellular transport and trafficking

DKF2phtes3\_17n18 encodes a novel 732 amino acid protein with weak partial similarity to known proteins.

The novel protein contains a ATP/GTP-binding site motif A (P-loop) and a TonB-dependent receptor protein signature 1. In *E. coli*, the tonB protein interacts with outer membrane receptor proteins that mediate uptake of specific substrates into the periplasmic space. In the absence of tonB these receptors bind their substrates but do not carry out active transport. The novel protein seems to be involved in ATP-dependent transport of substances into the cell.

The new protein can find application in modulation of cell-permeability and transport of suitable substrates into the cell.

unknown receptor

protein contains TONB\_DEPENDENT\_REC\_1 Pattern and ATP\_GTP\_A Pattern,

Sequenced by GBF

Locus: unknown

Insert length: 2853 bp  
Poly A stretch at pos. 2806, no polyadenylation signal found

```
1 GTCCTTTTAA GTCAGTAAAT TGAACAAAGT CGGTTATTGG GCAAGCAGTT
51 CCTATAAAAA ACTACATGGC TAAGGTTCTT AATGATTGAC CACAAGCAGA
101 TCTTTCACCC TCGGATCTCT AGCTACAAAA GGTCCCCACA CTGAAGAAGC
151 CACTACCTCC ACCACCAACA GCACCAACAC GTCCAGTGCT GCTGGCAACC
201 ACTGGGGCAG CCAAGCGGCT CACCCTCTCT CCCACCATGG CCCGTGAGGT
251 GCGCACCCAC CAGGAGACCC TGAACAGGTT TCAGCAGCAG TCCATCCACC
301 TGCTGACGGA GCTCCTCAGA CTGAAGATGA AGGCCATGGT GGAGTCTATG
351 TCGGTGGGTG CCAACCCCTT GGACATCACC AGGCCGTTTG TGGAGGCCAG
401 CCAGCTCCTC CACCTCAATG CCAAGGAGAT GGCCCTCAAC TGCCTGATCA
451 GCACAGCCGG GAGAAGTGGC TACAGCAGCG GACAGTTGTG GAAAGAGTCC
501 CTCGCAAAACA TGTCCGCCAT TCGGGTGAAC TCGCCCTTAC AGCTGATCTA
551 CCACTCTTCC ACAGGCTGTG TGAGCTTTTC TCTCTCTGCT GGAAAGAGAG
601 TCAAGAAGAA AATAGGCAAA TCTAGAACTA CAGAAAGATC CAGCATCCCG
651 CCCCTGCATC GAGGAGTGGG AACCCCTGCC AACAGGCTCG ACTTCAGCGA
701 CCCCTGCCCT GAGGCCCGGG AGAGCTGCGA GGAGTTGTGT CCCCACATAG
751 AGCTGAAAG GGCACATGG AACGGGAGGA ATATCTCTTA CCCCATGATC
801 TTACGAAACT ACAAGGCAAA GATGCCCTCT CATCTAATGT TGGCCCGCAA
851 AGGAGACTCT CAGACCCCGG GTTTACATTA CCCTCCACT GCAGGTGCTC
901 AGACTCTCAG CCCACCTCT CACCCTCTT CTGCCAACA TCATTTCAGT
951 CAGCATTGTC AAGAGGGGAA GGCACCCAAG AAGGCCCTCA AGTTTCATTA
1001 CACCTTCTAT GATGGCTCCT CCTTCGTTTA CTATCCCTCT GGAAACGTCG
1051 CTGTATGTCA GATCCCAACA TGCTGCAGAG GGAGAACCAT CACCTGCCTC
1101 TTTAATGACA TACCTGGATT CTCCTTGCTG GCCCTATTCA ATACTGAAGG
1151 CCAGGGCTGT GTTCACTACA ACCTAAGAAC CAGTTGCCCA TATGTCTTAA
1201 TCTTGGATGA GGAAGGTGGG ACCACCAATG ACCAGCAGGG CTATGTAGTC
1251 CACAAGTGGG GCTGGACTTC CAGGACAGAG ACCCTGCTTT CCCTGGAATA
1301 CAAGGTGAAT GAGGAAATGA AACTAAAGGT ACTGGGACAG GACTCCATCA
1351 CAGTCACCTT CACCTCCCTG AATGAGACAG TAACACTCAC TGTGTGGGCC
1401 AACAATTGTC CCCATGGAAT GGCATATGAC AAACGGCTGA ACCGCAGAAT
1451 CAGCAACATG GACGACAAGG TGTATAAGAT GAGCCGAGCC CTGGCTGAGA
1501 TCAAGAAGCG GTTTCAGAAG ACAGTGACTC AGTTCAATTA TTCTATCTTG
1551 CTGGCCGCGG GTCTGTTTAC CATTGAATAT CCCACCAAAA AGGAGGAGGA
1601 AGAATTTGTT CGGTTCAGA TGAGATCCAG AACTCATCCC GAGCGGCTCC
1651 CCAAGCTAAG TTTATACTCA GGAGAAAGTC TTTTACGATC TCAGTCAGGC
1701 CACCTGGAAT CCTCAATTGC AGAGACTTTG AAGGATGAGC CTGAGTCTGC
1751 TCCTGTGAGC CCAGTTCGGA AGACCACCAA AATCCACACC AAAGCCAAGG
1801 TCACATCCAG AGGGAAGGCC CGCGAGGGGC GCAGCCCCAC CAGGTGGGCG
1851 GCCTTGCCCT CAGACTGCCC GCTGGTGTCT CGGAAGCTCA TGTCTAAGGA
1901 AGACACCCGT GCTGGCTGCA AGTGCTGGT GAAGGCGCCC CTGGTCTCTG
1951 ACGTGGAGCT GGAGCGCTTC CTGTTGGCGC CCGAGAGCCC CAGCCAAGTG
2001 CTGGTGTGTT GGATCATCTC AAGCCAGAAC TACACCAGCA CTGGGCAGCT
2051 CCAGTGGCTG CTGAACACTC TCTACAACCA CCAGCAGCGG GGCCGTGGCT
2101 CCCCCTGCAT CCAAGTCCGG TATGACTCCT ACCGCTGCTG GCAGTATGAC
2151 CTGGACAGCC CCCTGCAGGA GGACCTCCCT CTGATGGTGA AGAAGAATCT
2201 TGTGGTGCAG GGGATGATTG TGATGTTTGC CGGGGGGAAG CTCATTTTGT
2251 GGGGCGGTGT TTTGAATGGA TATGGCTCTA GCAAGCAGAA TCTGCTGAAA
2301 CAGATCTTCC GGTCTCAACA GGATTACAAG ATGGGCTACT TCCTGCCGGA
2351 TGACTACAAA TTCAGTGTTC CCAACTTGTG CTTGAGCCTG GAGGATTCTG
2401 AATCAGTCAA GAAAGCCGAG TCAGAAGATA TCCAAGGAAG CAGGCTCTCA
```

WO 01/12659

PCT/IB00/01496

```
2451 TTGGCCCTGG AAGACTATGT GGAAGAAGGAG TTATCTCTGG AGGCTGAGAA
2501 GACAAGAGAG CCTGAAGTGG AGCTACATCC TCTCAGCAGG GACAGCAAGA
2551 TAACTAGTTG GAAGAAGCAG CCCTCAAGA AGTAGCGCCA TCCTGGCAGC
2601 AGCCAAGTGA GCCAGGCCCC GGGCCGGGGT GCTGGGGCTT CTGCCCAGCC
2651 CAGCCCTGCC TCCCCTGGTCT CCCACCTGT CCTCCAAGCT TCTATAATAA
2701 ACCAGCGGGC CTCCAGCATT GGGGTGAGGC TCTGGGAGAG GACAAAAAAA
2751 AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAGGGG
2801 CGGCCGAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAGGGG
2851 CCG
```

BLAST Results

No BLAST result

Medline entries

No Medline entry

Peptide information for frame 3

ORF from 237 bp to 2582 bp; peptide length: 782  
Category: putative protein  
Prosites motifs: ATP\_GTP\_A (122-130)  
TONB\_DEPENDENT\_REC\_1 (1-44)

```
1 MARQVRTHQE TLNRFQQQSI HLLTELLRLK MKAMVESMSV GANPLDITRR
51 FVEASQLLHL NAKEMAFNCL ISTAGRSYGS SQQLWKESLA NMSAIGVNSP
101 YQLIYHSSTA CLSFSLSAGK EAKKIKGSR TTEDVSMPLP HRGVGTPANS
151 LEFSDPCPEA REKLQELCRH IEAERATWKG RNISYPMILR NYKAKMPSHL
201 MLARKGDSQT PGLHYPTTAG AQTLSPTSHP SSANHHFSQH COEGKAPKKA
251 FKHYTFYDG SSFVYYPNGN VAVCQIPTCC RGRITITLFP DIPGFSLAL
301 FNTEGQGCWH YNLKTSCTPVV LILDEEGGTT NDQQGYVVHK WSWTSRTETL
351 LSLEYKVNEE MKLKVLGQDS ITVTFTSLNE TVTLTVSANN CPHGMAYDKR
401 LNRRISNMDD KVKYMSRALA EIKKRFQKTV TQFINSILLA AGFTIEYPT
451 KKEEEFVRFV KMRSRTHPER LPKLSLYSGE SLLRSQSGHL ESSIAETLKD
501 EPESAPVSPV RKTTKIHTKA KVTSRGKARE GRSPTRWAAL PSDCPLVLRK
551 LMLKEDTRAG CKCLVKAPLV SDVELERFLL APRDPSQVLV FGISSQNYT
601 STGQLQWLLN TLYNHQQRGR GSPCTQCRYS SYRLQYDLD SPLQEDPPLM
651 VKKNSVVQGM ILMFAGGKLI FGGRVLNGYG LSKQNLKQI FRSQDYKMG
701 YFLPDYKFS VPNSVLSLED SESVKKAESE DIQSSSSSLA LEDYVEKELS
751 LEAEKTRPE VELHPLSRDS KITSWKQAS KR
```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphtes3\_17n18, frame 3

No Alert BLASTP hits found

Pendant information for DKFZphtes3\_17n18, frame 3

Report for DKFZphtes3\_17n18.3

```
[LENGTH] 782
[MW] 88030.16
[pI] 9.22
[BLOCKS] BL00286 Squash family of serine protease inhibitors proteins
[PROSITE] ATP_GTP_A 1
[PROSITE] MYRISTYL 4
[PROSITE] CAMP_PHOSPHO_SITE 3
[PROSITE] CK2_PHOSPHO_SITE 14
[PROSITE] PROKAR_LIPOPROTEIN 1
[PROSITE] TONB_DEPENDENT_REC_1 1
[PROSITE] PKC_PHOSPHO_SITE 10
[PROSITE] ASN_GLYCOSYLATION 4
[KW] Alpha_Beta
```

**PCT/IB00/01496**

SEQ	MARQVTHQETLWRFQOQSIHLLELLRLKMKAMVESMSVGANPLDITRFRVEASQJLHLH
PRD	ccccchhh
SEQ	NKCHHAFNCLISTHAGRSGVSGQWKLSESLHMAHGVNPSYQOLIYHSSTACLFSFLSAGK
PRD	hh
SEQ	EAKKIKGSRITTEDEVSMPLPHRGVGT PANSLEFSDPCPEARKQLQELCRHIEAERATWKG
PRD	hh
SEQ	RNTISYPMILRNYKAKMPSHMLARKGDSQTPGLHYPTAGAQTLSPTSHSSANHHFSOH
PRD	ccccchhh
SEQ	CQEGAKPKKAFKFHYTFYDGSSFVYVPSGNVACQIPTCCRGRTITCLFNDIPGFSLLAL
PRD	ccccchhh
SEQ	FNTEGQGCVVHYNLKTSCPVYLIDEEGGTTNDQQGYVVKHWSWTSRTETLLSLEYKVNEE
PRD	eccc
SEQ	MKLKVLGQDSITVFTTSLNETVLTLSVANNCPHGMAKYOKRLNRRISNMDDKVKYMSRALH
PRD	hh
SEQ	EIKKRFRQKTVTQFINSILLAAGLTIEYPTKKEEEFVRFKMSRTHPERLPKLSLYSGE
PRD	hh
SEQ	SLERSQSGHLESSIAETHLKDEPESAPVSVPRKTKIHTKAKVTSRGKAREGRSPTRWAAL
PRD	eeeeccccchhh
SEQ	PSDCPLVLRLMLLKEDTRAGCKCLVAKPCLVSDVLEFLRLAPDPSQVLFVFGISSQNTT
PRD	ccccchhh
SEQ	STQGLQMLLNTLHYNQQRGCSPCIQCRYSYRLQLYDLSPLEQDPPLVMKXNSVVGQM
PRD	ccccchhh
SEQ	LMFAGGKLLFGGRVLNHYGLSKQNLQLQIFRSQQQYKMGVFLPDDYFVSVPNSVLSLED
PRD	heeeeeccccccccccccccccchhh
SEQ	SESVKKAESDEDQSGSSSLALEDYVEKELSLEAKETREPEVELHPLSRDSKITSWKKQAS
PRD	chhh
SEQ	KK
PRD	cc

Prosites for DKFZphtes3\_17n18.3

PS000001	91->95	ASN GLYCOSYLATION	PDCC000001
PS000001	182->186	ASN GLYCOSYLATION	PDCC000001
PS000001	379->383	ASN GLYCOSYLATION	PDCC000001
PS000001	598->602	ASN GLYCOSYLATION	PDCC000001
PS000004	403->407	CAMP_PHOSPHO_SITE	PDCC000004
PS000004	511->515	CAMP_PHOSPHO_SITE	PDCC000004
PS000004	652->656	CAMP_PHOSPHO_SITE	PDCC000004
PS000005	48->51	PKC_PHOSPHO_SITE	PDCC000005
PS000005	177->180	PKC_PHOSPHO_SITE	PDCC000005
PS000005	344->347	PKC_PHOSPHO_SITE	PDCC000005
PS000005	450->453	PKC_PHOSPHO_SITE	PDCC000005
PS000005	497->500	PKC_PHOSPHO_SITE	PDCC000005
PS000005	513->516	PKC_PHOSPHO_SITE	PDCC000005
PS000005	523->526	PKC_PHOSPHO_SITE	PDCC000005
PS000005	631->634	PKC_PHOSPHO_SITE	PDCC000005
PS000005	723->726	PKC_PHOSPHO_SITE	PDCC000005
PS000005	774->777	PKC_PHOSPHO_SITE	PDCC000005
PS000006	7->11	CK2_PHOSPHO_SITE	PDCC000006
PS000006	131->135	CK2_PHOSPHO_SITE	PDCC000006
PS000006	256->260	CK2_PHOSPHO_SITE	PDCC000006
PS000006	329->333	CK2_PHOSPHO_SITE	PDCC000006
PS000006	345->349	CK2_PHOSPHO_SITE	PDCC000006
PS000006	377->381	CK2_PHOSPHO_SITE	PDCC000006
PS000006	406->410	CK2_PHOSPHO_SITE	PDCC000006
PS000006	450->454	CK2_PHOSPHO_SITE	PDCC000006
PS000006	466->470	CK2_PHOSPHO_SITE	PDCC000006
PS000006	493->497	CK2_PHOSPHO_SITE	PDCC000006
PS000006	497->501	CK2_PHOSPHO_SITE	PDCC000006
PS000006	571->575	CK2_PHOSPHO_SITE	PDCC000006
PS000006	693->697	CK2_PHOSPHO_SITE	PDCC000006
PS000006	717->721	CK2_PHOSPHO_SITE	PDCC000006
PS000008	145->151	MYRISTYL	PDCC000008
PS000008	327->333	MYRISTYL	PDCC000008
PS000008	592->598	MYRISTYL	PDCC000008
PS000008	734->740	MYRISTYL	PDCC000008

WO 01/12659

PCT/IB00/01496

PS00013	101->112	PROKAR_LIPOPROTEIN	PDOC00013
PS00017	122->130	ATP_GTP_A	PDOC00017
PS00430	1->44	TONB_DEPENDENT_REC_1	PDOC00354

(No Pfam data available for DKF2phtes3\_17n18.3)

DKF2phtes3\_18f3  
-----

group: testes derived

DKF2phtes3\_18f3 encodes a novel 248 amino acid protein with partial similarity to human TNF-inducible protein CG12-1.

The novel protein contains two leucine zippers.

No informative BLAST results; No predictive prosite, pfam or SCOP motif.

The new protein can find application in studying the expression profile of testis-specific genes.

similarity to TNF-inducible protein CG12-1

Sequenced by MediGenomix

Locus: unknown

Insert length: 4608 bp

Poly A stretch at pos. 4570, polyadenylation signal at pos. 4550

```
1 GACAGAAATG AATGGGAATG GAGAGGCCGG CGGCCCGGGA GCCGCATGGG
51 CCCGACGCGC TGCAGCGCTT CCAGGGACTG CTGCTGGACC GCCGAGGCCG
101 GCTGCACCGC CAGGTGCTGC GCCTGCGCGA GGTGGCCCGG CGCTGGAGC
151 GCCTGCGCAG GCGCTCCCTC GTAGCCAACG TGGCCGCGAG CTGCTGAGC
201 GCAACGGGCG CCCTCGCCGC CATCGTGGGG CTCTCGCTCA GCCCGGTAC
251 CCTGGGGACC TCGCTGCTGG TGTGGGCGGT GGGGCTGGGG GTGGCCACAG
301 CCGGAGGGGC CGTCACCATC ACGTCCGATC TCTCGCTGAT CTCTGCAAC
351 TCCCGGGAGC TCGGAGGGGT GCAGGAGATC GCGGCCACCT GCCAGGACCA
401 GATGCGAGAG ATCCTGAGCT GCCTCGAGTT TTTCTGCCGC TGGCAGGGCT
451 GCGGGGACCG CCAGCTGCTG CAGTGGCGGA GGAACGCCCT CATCGCCCTG
501 TACAATTCTG TCTACTTCAT CGTCTTCTTT GGCTCACGTG GCTTCTTCAT
551 CCCAGGCGCG GCGGAGGGGG ACACCAAGGT TAGCCAGGCC GTGCTGAAGG
601 CCAAGATTCA GAAACTGGCC GAGAGCCTGG AGTCTGACAC CGGGGCTCTG
651 GACGAATCA GCGAGCAGCT GGAGTCTCGG GTTCAGCTCT GCACCAAGTC
701 CAGTCGTGGC CAGGACCTCA AGATCTCTGC TGACCAGCGT GCAGGGCTGT
751 TTTTCTGAGA ACATCCTTTC CCCCTAATGA CCGAGGCCAG CAATCATCC
801 TCATGGGATG CTCAGAAATT TGTAAGTCCC TTAGGAAAAC ACCAAGCTGG
851 GTTAGGAGCC GAAGGCAAAAG GATGAGAAAA ACTGTTTTTG AAGTGGGCG
901 GTCCCCAAAG CCCTTCTTTT CCCATCACTG TGACATCTGC CTGGGCTTGA
951 GTGCTACGGA CTTTTCAGTC TTCCTAGTGG AAAAATGTGA CCAAAAACCT
1001 CCTTTTCTTT TATCAAAAAC TTCTGTCTTA AACACAGCTG GGCAGGCAC
1051 CCTGTTTTAA AGTTATTTTC GGGTCCCTGA CCCTGCCCTG GTGGCTTGGC
1101 CTGAGACTGG AGAGAGTGCC ATCCTCTGGG TCCTCTCCAA GTCTACTAG
1151 TCTTTGAAGT CCTCAAAATG TGGGTGAGGA AGGCATTTGC CTCTATTCCA
1201 GAAATTTCTG TACAAAGAAC TCCAGAAATC GAGGCAAAAT AGCCCTTCTC
1251 TGAACGTTGT AGGATGGTTC AGAACCCAGA GAGGCAAAAT GTGCTGATAT
1301 CTCCTCTCTT TCCCTTTCCC CTCAGCTTAC TTACTCCCAG ATGCGGCCCTG
1351 GGTATGAAGT AGGCTTTTCC TGACTGGCTC CCAATCCAGT CCTCAAAGTA
1401 CTCAGAGGGG AAGCCCGTGA AGCCGTCTAT TAAGTCTCTG TCCTCACAT
1451 GAAGCTGAGG CCGAGATAGA TCGAGGCACT GCCAACTTCA TTCCCGACA
1501 TCATTGTGTT CAGAAGAGAG TGATGGGTTT TGAGTACAGC AGTCTGGGC
1551 TTGAGACAGG CTTTGTCACT ACTGTGTGAG TGTAGCCACC TAATCTCTCT
1601 GAGACTGTGT AAAACAAAGA TGATAAAATC TCACCTGTGT GTGAGATATT
1651 AAATGAGCCA AAGTGCCTAG CATGATGGTG CTGGCTCATA TACTGTAGTC
1701 CCTGGAATGG CAAATTAACA TCACCCAGGA ACTTGTTAGA AAGGCAAAAT
1751 CTTGGACACA ACCCTCCTGA TTTATGGAAT CAGAACTCT GGCTGTGGGG
1801 CCCAGCAACC TGAGTTTAAA CAATTTCTCT GGGTGGTTCT GCGGCACACT
1851 AAGGTTTGAA AATCACTACA ACAAATGCTA ACTTCTAATC CCCTTGATGA
1901 GCTTTCACGA AGTCTCACGG CTCTCTAGG GACTCCATGG TCTTCAGAGT
1951 CGTTCACAGA TGACCAAGGA CAGACTGTGT CCCAGAAGCC AAAATGAGAG
2001 AGAGAGAGAG AGCAGCGGTA CGTGACCCCT GGGGAGTGT CTACCGGTAT
2051 GAATAAGGGA TGTAACTACT AAAGCCCATT AGGGGGCAGT GTTTCGCGCC
2101 TGTGTAGAAA ACTGGTACAG AAAGGATCCT ATATGAAGTT CCTGAAACTG
2151 ACCTTTGTCT ATTATTACCT TCTCTGAAAA GTGCCAGTCC ATGTATTTTT
2201 TATTATTTTT AAGTTTGTAA TTTAATTTTT AATTATTGTT TAGTGTTTGC
2251 ATTTAATTTT ATTTAATCAC CACATTTAGA AAATAATAAG AGCAAGTTTC
2301 TAAATGGGAG ACTGCTGAGG CTCTTTGCAA GAGATGAGAT TAAGTTTGAG
2351 TTTCTAAGGC AGGGCATGAG CTGGAATAG CATTGCTTTC CTTGATTGTC
2401 TCTCTCCTTC AGGGAGATTC TTTTCTCTTA GTGTTTTAAG TGATCCTTTG
2451 AAGTAAGTGT GGAGAGTCTT AATGGCAAG ACCAGGAGCT GAGTTTAAGC
2501 TTGTAATGGA AGCTTGCAAT GTGGGATATA TAACTGAGGA AGCATATTTA
2551 TCCTGAAGGT ATTTTGCCAG AAGGTATCAC TTGACCTGGA AAAGGAATCT
2601 ATTTAGTTCA GGAAGATAAA AAGTTTAGA GGTATGTGAA GGAAGCACTT
2651 AGAACTTGCA AGCCTGATGT CCTATCAAGT TATGTCTTCT GGGTGACAGA
2701 CAAAATAGCT TGCTTATGCG TGGTATGATG TTGCATTTTC AGTTTGGGGT
```



```

2751 CTGTAGAAA CTGTCAGTGA AAATATGTAC AATTCCTTCA ATTTCCATTG
2801 TTAACAACTG TAATGTTGAA AAATAAGTTG AAAAGTCTTT GGGACCATAC
2851 ATGCAAAAAC GGTGCTCTG TTAATAATT ATTTAATATT CTATAAATGT
2901 ACCCAATCTG TCCGCACCTT TCCAGTGAT GGGGAGTAT GTCTGAGGAA
2951 GTATAATTTC AGTACTGGGG TCGGGAGAG GAGGTGATGT TTCTACATTT
3001 TTATTTTTTC TATAAATTGC AATTGGCTGT TATGCTGGTT TATTTTGAAA
3051 TTTATATTGG TTTCTTTTCA AGCTGGTGTG ATCTCCTAGA CTGTTTCACC
3101 CAGATGCTAG CATTTTTTTT TTTTGTAGA CAGAGTCTCA CTCTGTCACC
3151 TAGGCTGGAG TTGCACTGGT TTGATCTCGG CTCAGTGCAG CCTCCGACTC
3201 CTGGGTTCAG GCAATTCTTC TGCTCAGCG TCCTGAGTAG CTGGGATTAC
3251 AGATGTGCAC CAGCACACCC GGCTAATTTT TTGTATTTT AGTAGAGACA
3301 GGGTTTCGCC ATGTGGGCCA GGCTGGTCTT GAATCCTGG CCTTATGTGA
3351 TCCGCCACCC TTGGCTTCCC AAAGTGTGGG GATTACAGGC ATGAGCCACC
3401 TCGCTGGGCC AGATGCTAGC ATTTTAGATC AAACAATTCA TTTTAGATGA
3451 ATGTGTTTGT TTCACAATCA TTTTAAATCA TTTTGAATG TACTTCACAT
3501 TATTAGTTGT GTTATGGCAT AAAGGTACAA CCATTCCCTA ACTCCATCTT
3551 TTATTAATGC TTAAGTTTAA ATTATATTCT TCCAATGCCT AAGCTATTCC
3601 CTAGAATTAA ACTGGGCACT TTTGGAAGCA GCAACAGTAA CAGCAGCAGC
3651 AAACTTTTC TCTCATATTT TGGGTGTATC AAAAGTTCTA GACTTTTGAA
3701 GTTATGATTG CAGTGGCCCA CTTTATTTCT AAGGAAGAGT GTCTACTTTG
3751 GAACGATACT TTGCACATAG TAGGAATCTA AGAAATACAT TTGAATAATT
3801 ATAATTAAT GTTTAGCTAT CTTAATGAGA ATTTGTTGAC AACAAAAGAT
3851 CATCCATCGC CTTATGTGTG AGTAAGATTG GAGCCTCTAT CAAGATTTAG
3901 TCAAGTTTCAG TTAGATTGAT TCTAGAAACA AATATTATT TCTTTCTTTT
3951 ACGGGGATGT GAATAAGGCT TTTCTTAAG GCCTTCATTC TTTAAACAAA
4001 CAGGTTGAAA TGGTATGTTG TAAAGAGAGA GACGGGAGAG AGGTATTTAG
4051 ATGATAAGTG TACTTCACAA AAATGCCAAA GTTTGAAAAA TAGGTATGTT
4101 TGTCTAAAT GTTTAAGTGC TTCTCTGTTA GGTCTGGGG CTGCAATCA
4151 TTTGAATTGT TCTGTTTAC AATAAAGGAG ATTCAGTGGG TTCTGCATTT
4201 TCAGGATTCA ATAGAACTGC TCCATTAATA AATAATCCCT TAGCAAGCAT
4251 TCGAATCCTA ACTGCTTTGA TGCATTGCC CTCGGGCACC TGTCATTTC
4301 AATATGGTAG GTGTCAAAGT CAAAAGTATT TACTGGGAGA AAAAGAGAG
4351 GAGTGGTTGT AGAAGTCTCC CTAATCAGA CATGTCAAGC AATCAGCCAA
4401 CGTGGTGTAT TTCTCATTCA ATATTTTAGT GTGAATTGAG ACACAGAGAT
4451 AAAGACATCG TGCAGAGATA AATGGGGATA CAGTTAAATG TAGCAACTCT
4501 TGAGTTCAAT TTTTCCCACT GTAGCAAAAT TAATGCTTTC TCTTTATTGA
4551 AATAAATTGC TCATTCCCTC AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA
4601 AAAAAAGG

```

## BLAST Results

Entry HSG27587 from database EMBL:  
human STS SHGC-32548.  
Score = 1951, P = 9.0e-101, identities = 411/425

Entry HS073350 from database EMBL:  
human STS EST303564.  
Score = 1417, P = 8.7e-58, identities = 285/287

## Medline entries

No Medline entry

## Peptide information for frame 2

ORF from the beginning to 580 bp; peptide length: 194  
Category: questionable ORF  
Classification: no clue

## BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphtes3\_18f3, frame 2

PIR:CGBOIS collagen alpha 1(I) chain - bovine (fragments), N = 1, Score = 155, P = 4.5e-10

TREMBL:HSCG1PA1\_1 gene: "COL1A1": Human proalpha 1 (I) chain of type I procollagen mRNA (partial), N = 1, Score = 155, P = 6.5e-10

>PIR:CGBO1S collagen alpha 1(I) chain - bovine (fragments)  
Length = 779

## HSPs:

Score = 155 (23.3 bits), Expect = 4.5e-10, P = 4.5e-10  
Identities = 60/152 (39%), Positives = 67/152 (44%)

Query: 7 GEAGGPGAAWARRAAALPGTAA--GPPRPAAPPGA--APARGGPAPGAPAQALPRSQRG 62  
G+ G PG + AR PG GPP PA P GA AP G A A P SQ  
Sbjct: 230 GDLGAPGPGSGARGERGFPGERGVEGPPGAPRGANGAPGNDGAKGDAGAPGAPGSGQAP 289

Query: 63 QLAERNRGRPRRHGALAQPQHPGDLAAGVGRGAGGGHSRRGRHHVRSADLLQLPGAEE 122  
L G P RGA PG GD +GA G + G VR L + PG A  
Sbjct: 290 GL---QGMPGE-RGAAGLPGPKGDRGDAGPKGADGAPGKDG---VRGLTGPIGPPGPAG 341

Query: 123 GAGDRGHL-P-GP-----DARDPELPRVFLPLAGLRGPPAA 156  
GD+G P GP D +P P P AG GPP A  
Sbjct: 342 APGDKGEAGPSGPAGTRGAPGDRGEPGPPG---P-AGFAGPPGA 381

Score = 121 (18.2 bits), Expect = 5.4e-05, P = 5.4e-05  
Identities = 52/154 (33%), Positives = 60/154 (38%)

Query: 7 GEAGGPGAAWARRAAALPGTAAGPPRPAAPPGAAPARG----GPAPGAPAQALPRSQRG 61  
G G PGAA R P AGPP P P G ++G GPA G P + P G  
Sbjct: 434 GATGFPGAA-GRVGPFGPSGNAGPPGPPGAGKEGSKGPRGETGPA-GRPGEVGPFGPPG 491

Query: 62 RQLAERNRGRPRRHGALAQPQHPGDLAAGVGRGAGGGHSRRGRHHVRSADLLQLPGA 121  
A G P G PG PG RG G +RG R L PG +  
Sbjct: 492 P--AGEKGAPGAD-GPAGAPCTPGQGIAGQGVVGLPGQRGE---RGFPGL---PGPS 541

Query: 122 EGAGDRGHLPGPDARDPELPRVFLPLAGLRGPPAAVRE 160  
G +G R P P + GL GPP + RE  
Sbjct: 542 GEPGKQGPSGASGERGPPG---MGPPGLAGPPGESGRE 577

Score = 117 (17.6 bits), Expect = 1.8e-04, P = 1.8e-04  
Identities = 52/148 (35%), Positives = 62/148 (41%)

Query: 7 GEAGGPGAAWARRAAALPGTAAGPPRPA---PPGAAPARGGPAPGAPAQALPRSQRG-R 62  
G G PG AR +A PG A G P A PPG + GP PG P A +G R  
Sbjct: 416 GNVGAPGPKGARGSAGPPG-ATFFPGAAGRVGPPGPS-GNAGP-PGPPGAGKEGSKGPR 472

Query: 63 QLAERNRGRPRRHGALAQPQHPGDLAAGVGRGAGGGHSRRGRH--HHVRSADLLQLPGA 120  
GRP G + PG FG GA G + ++ LPG  
Sbjct: 473 GETGPAGRP---GEVGPFGPPGAGEKAGADGAPAGAPGTGPGQGIAGQGVVGLPGQ 528

Query: 121 AEGAGDRGH--LPGPDARDPEL-PRVFLPLAGLRGPP 154  
G+RG LPGP + P +G RGPP  
Sbjct: 529 R---GERGFPLPGPSGEPGKQGPS---GASGERGPP 559

Score = 117 (17.6 bits), Expect = 1.8e-04, P = 1.8e-04  
Identities = 54/162 (33%), Positives = 64/162 (39%)

Query: 7 GEAGGPGAAWARRAAALPGT--AAGPPRPAAPPGAAPARG--GPA--PGAPAQALPRSQR 60  
G G PG + PG A+GP P PPG G G A PG P + P +  
Sbjct: 29 GPPGAPGPGGFGQPPGEPGEGASGPMGPRGPPGPKNGDDGAGKPRPGERGPPGPQ 88

Query: 61 G-RQLAERNRGR--RRRHGALAQPQHPGDLAAGVGRGAGGGHSRRGRHHV--RSLADLL 115  
G R L G P + HRG G GD +G G G + R L  
Sbjct: 89 GARGLPCTAGLPGMKHGRGFSGLDGAAGDAGPAGPKGEPGSPGENGAPGQMGPRGLPGFP 148

Query: 116 QLPGAA--EG-AGDRGHLPGPDARDPELPRVFLPLAGLRGPPAAA 157  
GAA G AG+RG +PGP P AG +GPP A  
Sbjct: 149 GPKGAAGEPKKAGERG-VPGPPGAVG--PAGKDCEAGAQGPPGPA 190

Score = 113 (17.0 bits), Expect = 5.4e-04, P = 5.4e-04  
Identities = 54/148 (36%), Positives = 58/148 (39%)

Query: 7 GEAGGPGAAWARRAAALPCTA----AGPPRPAAP---PGAAPARGGPAP-GAPAQALPR 57  
G AG PGA A PG A AGPP PA P PG G P P GA A P  
Sbjct: 374 GFAGPPGADGQPGAKGEPGDAGAKGDAGPPGAPGAPGPGPIGNVGPAGPKGARGSAGPP 433

Query: 58 SQRGRQLAERNRGRPRRHGALAQPQHPGDLAAGVGRGAGGGHSRRGRHHVRSADLLQL 117  
G A P G PG PG +G G GR V  
Sbjct: 434 GATGFPGAAGRVGPPGPSNAGPPGPPGAGKEGSKGPRGETGPAGRPGEVGP----- 486

Query: 118 PGAAEGAGDRGHLPGPD--ARDPELPRVFLPLAGLRG 152  
PG AG+G PG D A P P +AG RG  
Sbjct: 487 PGPPGPAGEKG-APGADGAPAGPTGPP-QGIAGQRG 521

Score = 110 (16.5 bits), Expect = 1.3e-03, P = 1.2e-03

Identities = 54/151 (35%), Positives = 60/151 (39%)

Query: 7 GEAGGPGAARAAALPGTAAGPPRPAAPPG--AAPAR-GGPAP-GAPALPRSRGR 62  
GE G G A + L P G A G P P A P G P G P P G A + +RG  
Sbjct: 194 GERGEQGPAGSPGFQGLPGPA-GPPGEAGKPGEGVPGDLGAPGSGARGERGFPGERV 252

Query: 63 QLAERNRPRRRHRLAQPQHPGDLAAGVGRGAGGGHSRRGRHHHVRSLADLLQLPGAEE 122  
+ PR GA G GD A G+G +G R A L PG  
Sbjct: 253 EGPPGPAGPRGANGAPGNDGAKGDAGAPGSGQAPGLQGMPE-RGAAGL---PGPK- 307

Query: 123 GAGDRGHLPGPDARD--PELPRVFLPLAGLRGPPAAA 157  
GDRG GP D P V L G GPP A  
Sbjct: 308 --GDRGDA-GPKGADGAPGKDTV-RGLTGPIGPPGPA 340

Score = 109 (16.4 bits), Expect = 1.7e-03, P = 1.7e-03  
Identities = 55/154 (35%), Positives = 60/154 (38%)

Query: 4 NGN-GEAGGPGAARAAALPGTAAGPPRPAAPPGAAPARG-GPAPGAPALPRSRGR 61  
NG+ GEAG PG R P A G P A PG RG GA A P +G  
Sbjct: 67 NGDDGEAGKPRP-GERGPPGPQARGLPGTAGLPGMKHGRGFSGLDGAKGDAGPAGPKG 125

Query: 62 RQLAE-RNGRPRRRHRLAQPQHPGDLAAGVGRGAGGGHSRRGRHHHVRSL----ADLL 115  
+ NG P + G PG PG A G G G V A  
Sbjct: 126 EPSPGENGAPGQ-MGPRGLPGFPKGAAEPKAGERGVPVPGGAVGPAGKDGEAGA 184

Query: 116 QLPGAAGAGDRGHLPGPDARDPELPRVFLPLAGLRGPPAAA 157  
PG A AG+RG GP A P F L G GPP A  
Sbjct: 185 GPPGPAGPAGERGE-QGP-AGSPG----FQGLPGPAGPPGEA 220

Score = 104 (15.6 bits), Expect = 6.6e-03, P = 6.6e-03  
Identities = 44/131 (33%), Positives = 49/131 (37%)

Query: 2 EVNGNGEAGGPGAARAAALPGTAAGPPRPAAPPGAAPARGGPAP-GAPALPRSRGR 60  
E GE G PG R L P G GP A PG A RG P P G A A +  
Sbjct: 126 EPSPGENGAPGQMGPR--GLPGFP-GPKGAAGEPKAGERGVPVPGGAVGPAGKDGEA 181

Query: 61 GRQLAERNRPRRRHRLAQPQHPGDLAAGVGRGAGGGHSRRGRHHHVRSLADLLQLPGA 120  
G Q P PG G PG G+ G G + DL PG  
Sbjct: 182 GAQGPAGPAGERGEQGPAGSPG--FQGLP-GPAGPPGEAGKPGEGVPGDL-GAPGP 237

Query: 121 AEGAGDRGHLPG 132  
+ G+RG PG  
Sbjct: 238 SGARGERG-FPG 248

Score = 104 (15.6 bits), Expect = 6.6e-03, P = 6.6e-03  
Identities = 43/131 (32%), Positives = 55/131 (41%)

Query: 7 GEAGGPGAARAAALPGTAAGPPRPAAPPGAAPARGGPAPGAPALPRSRGRQLAE 66  
GEAG G A R A PG G P P P G A GP PGA Q + + G A +  
Sbjct: 347 GEAGSPGAGTRGA---PGDR-GEPPGPAGFA---GP-PGADGQPGAKGPGDAGAK 397

Query: 67 RNRPRRRHRLAQPQHPGDLAAGVGRGAGGGHSRRGRHHHVRSLADLLQLPGAEGAGD 126  
+ P G PG G++ A +GA G G + A + PG + AG  
Sbjct: 398 GDAGPPGPAGPPGPIGNVGAPGPKARGSGATGFPGA-AGRVGPPGSGNAGP 456

Query: 127 RGHLPDPDARD 137  
G PGP ++  
Sbjct: 457 PGP-PGPAGKE 466

Score = 104 (15.6 bits), Expect = 6.6e-03, P = 6.6e-03  
Identities = 56/162 (34%), Positives = 62/162 (38%)

Query: 7 GEAGGPGAARAAALPGTAA--GPPRPAAPPGAAPARGGPAPGAPALPRSRGRQL 64  
G G PGA A G GP P P G A ARG P P Q PR +G  
Sbjct: 608 GPPGAPGAPGVPAGKSGDRGETGPAGPIGPVGPAGARG---PAGP-QG-PRGBKGZTG 662

Query: 65 AERNRPRRRHRLAQPQHPGDLAAGVGRGAGGGHSRRGRHHHVRSLA-DLLQ-LPG 119  
+ + HRG PG PG GA G RG S D L LPG  
Sbjct: 663 ZZGBRGIKHGRGFSCLQPPGPGSPGEQGPSGASGPAGPRGPPSAGSPGKDLNGLPG 722

Query: 120 AEGAGDRGHL--PGPDARDPELPRVFLPLAGLRGPPAAVREERLHRFPVQ 168  
G RG GP A P P P G GPP+ L +P Q  
Sbjct: 723 PIGPPGPRGRTDAGP-AGPPGPPG---P-PGPPGPPSGGYDLSFLPQPPQ 768

Score = 101 (15.2 bits), Expect = 1.5e-02, P = 1.5e-02  
Identities = 49/148 (33%), Positives = 55/148 (37%)

Query: 7 GEAGGPGAARAAALPGTAAGPPRPAAPPGAAPARGGPAPGAPALPRSRGR 62  
G AG PG A R PG A GP A G A+G P P PA + P G  
Sbjct: 152 GAAGEPGKAGERGVPVPGPPG-AVG---AGKDGEAGAQPPGPAGPAGERGEQGPAGSPGF 207

Query: 63 QLAERNRPRRRHGALAPGHGPDLAAGVGRGAGGGHSRRGRHHVRSADLLQLPGAEE 122  
Q P G + G PGDL A G G RG R + PG A  
Sbjct: 208 QGLPGPAGPPGEAGKPGEGVPGDLGAP---GPSGARGERGFPGE-RGVEGP---PGFAG 260

Query: 123 GAGDRGHLPGPDARDPELPRVFLPLAGLRGPP 154  
G G PG D + P G +G P  
Sbjct: 261 PRGANG-APGNDGAKGDAGAPGAP--GSQGAP 289

Score = 100 (15.0 bits), Expect = 1.9e-02, P = 1.9e-02  
Identities = 40/130 (30%), Positives = 48/130 (36%)

Query: 7 GEAGGPGAARAAALPGT--AAGPPRPAAPPGAAPARG--GPA--PGAPAQALPRSQR 60  
G G PG + PG A+GP P PFG G G A PG P + P +  
Sbjct: 29 GPPGAPGPGQGPGEPGEPEGASGPMGPRGPPGPGKNGDDGAGKPGRPGERGPPGPGQ 88

Query: 61 G-RQLAERNRPR--RRHRGALAPGHGPDLAAGVGRGAGGGHSRRGRHHVRSADLLQL 117  
G R L G P + HRG G GD +G G G + L  
Sbjct: 89 GARGLPGTAGLPGMKHGRGFSGLDGAKGDAGPAGPKGEPGSPGENGAPGQMGPRG-LPGF 147

Query: 118 PGAAEGAGDRG 128  
PG AG+ G  
Sbjct: 148 PGPKGAAGEPG 158

Score = 99 (14.9 bits), Expect = 2.5e-02, P = 2.5e-02  
Identities = 53/156 (33%), Positives = 61/156 (39%)

Query: 7 GEAGGPGAARAAALPGT--AAGPPRPAAPPGAAPARG--GPA--PGAPAQAL 55  
G G PGA R A PG A G P P P G + RG GPA P PA A  
Sbjct: 587 GRDGSFGAKDRGETGFAGAPGPPGAPGAPGVGPAKSGDRGETGPAGPIGPVGPAGAR 646

Query: 56 -----PRSQGRQLAERNRPRRRHGALAPGHGPDLAAGVGRGAGGGHSRRGRHHV 108  
PR +G + + + HRG G PG + +G G G  
Sbjct: 647 GPAGPQGPBGKGTGZZGBRGIKGRGFSGLQGPPGPPGSGEQGSGASGPAGPRGP- 705

Query: 109 RSLADLLQLPGAEEGAGDRG--HLPGPDARDPELPRVFLPLAGLRGPP 154  
PG+A G G LGGP P PR AG GPP  
Sbjct: 706 -----PGSAGSPKDGGLNGLPGPIG--PPGPRGRTGDAGPAGPP 742

Score = 98 (14.7 bits), Expect = 3.3e-02, P = 3.3e-02  
Identities = 51/158 (32%), Positives = 58/158 (36%)

Query: 7 GEAGGPGAARAAALPGTAA----AGPPRPAAPPGAAPARGGAP-GAPAQALPRSQR 60  
G G G R AA LPG AGP PG RG P G P A +  
Sbjct: 287 GAGPLQGMGERGAAGLPGPKGDRGDAGPKGADGAPGKDGVRGLTGPIGPPGAPAGDK 346

Query: 61 GRQLAERNRPRRRHGA---LAQPGHGPDLAAGVGRGAGGGHSRRGRHHVRSADLLQL 117  
G A +G P RGA +PG PG GA G + D  
Sbjct: 347 GE--AGPSG-PAGTRGAPGRGEPGPPGAGFAGPPGADQPGAKGEPGDAGAKGDAGP- 402

Query: 118 PGAAEGAGDRGHLPGPDARDPELPRVFLPLAGLRGPPAAVR 159  
PG A AG G + A P+ R G G P AA R  
Sbjct: 403 GPGAPGAPGPPGIGNVGAPGPKGARGSAGPPGATGFPGAAGR 444

Score = 96 (14.4 bits), Expect = 5.7e-02, P = 5.5e-02  
Identities = 46/152 (30%), Positives = 57/152 (37%)

Query: 6 NGEAGGPGAARAAALPGTAA--GPPRPAAPPGAAPARGGAPGAPGAPA-QALPRSQRGR 62  
+G G PGA + PG G PA PG A G P P PA ++ R + G  
Sbjct: 574 SREGAPGAEGSPGRDGSFGAKGDRGETGPAGAPGPPGAPGAPGVGPAKSGDRGETGP 633

Query: 63 QLAERNRPRRRHGALAPGHGPDLAAGVGRGAGGGHSRRGRHHVRSADLLQLPGAEE 122  
P RG G G+ +G G RG HR + L PG  
Sbjct: 634 AGPIGPVGPAGARGPAGPQGPGRB-----KGZTGZZGBRGIKGRGFSGLQGPPGPPG 686

Query: 123 GAGDRGHLPGPDARDPELPRVFLPLAGLRGPPAAA 157  
G+G P A P AG RGPP +A  
Sbjct: 687 SPGEQG--PS-GASGP-----AGPRGPPGSA 709

Score = 94 (14.1 bits), Expect = 9.7e-02, P = 9.2e-02  
Identities = 45/134 (33%), Positives = 56/134 (41%)

Query: 24 PGTAAGPPRPAAPPGAAPARGGPA-PGAPAQALPRSQRGRQLAERNRPRRRH--GALAQ 80  
P G P P PG +G P PG P + P RG G P ++ G +  
Sbjct: 21 PSGPRGLPGPPGAPGPGQGFQGPPEGEPEGASGPMGPRGPP-----GPPKNGDDGAGK 75

Query: 81 PGHPGDLAA-GV--GRGAGGGHSRRGRHHVRSADLLQLPGAEEGAGDRGH--LPGPDA 135  
PG PG+ G RG G G HR + L G A AG +G PG +  
Sbjct: 76 GPRPGERGPPGQARGLPGTAGLPGMKGH-RGFSGLDGAKGDAGPAGPKGEPGSPGENG 134

Query: 136 RDELP-PRVFLPLAGLRGPPAAA 157  
++ PR LP G GP AA

Sbjct: 135 APGQMGPGRG-LP--GFPGPKGAA 154

Score = 92 (13.8 bits), Expect = 1.7e-01, P = 1.5e-01  
Identities = 52/155 (33%), Positives = 58/155 (37%)

Query: 7 GEAGGPGAAWARRAAALPGTAAGPPRPAAPPGAAPARGGP-APGAPAQALPRSQRGRLA 65  
GEAG G A R A G GPP PA G A G P A G P A + G  
Sbjct: 347 GEAGPSGPAGTRGAPGDRGEP-GPPGPAFAGPPGADGQPGAKGEPGDAGAKGDAGPPGP 405

Query: 66 ERNGRPRRRHREGALAQPGHPGDLAAGVGRGAGGGHRRGR--HHHVRSLADLLQLPGA-- 121  
P G + PG G + GA G GR A PG A  
Sbjct: 406 AGPAGPPGPIGNVGAAPGPKGARGAGPPGATGFPGAAGRVGPPGPGSNAGPPGPPGPAGK 465

Query: 122 EGA-GDRGHLPGPDARDPELPRVFLP-LAGLRGPPAA 156  
EG+ G RG GP R E+ P AG +G P A  
Sbjct: 466 EGSKGPRGET-GPAGRPGEVGP GPPGPAKEGAPGA 501

Score = 92 (13.8 bits), Expect = 1.7e-01, P = 1.5e-01  
Identities = 51/156 (32%), Positives = 57/156 (36%)

Query: 7 GEAGGPGAAWARRA---AALPGT--AAGPPRPAAPPGAAPARGGPAPGAPAQAL-PRSQR 60  
G G PGA R A PG A G P P P G + RG P P + P R  
Sbjct: 587 GRDGSAGAKDRGETGFAGAPGPPGAPGAPGVPAGKSGDRGETGFAGPIGVGPAGAR 646

Query: 61 GRQLAERNRPRRRHREGALAQPGHPGDLA-AGVG--RGAGGGHRRGRH--HHVRSLADLL 115  
G A G PR +G + G G G +G G G A  
Sbjct: 647 GP--AGPQG-PRBGKZTGZGZBGRIKGRGFSGLQGGPPGPGSGEQGPGSGASGPAGPR 703

Query: 116 QLPGAAGAGDRG--HLPGPDARDPELPRVFLPLAGLRGPP 154  
PG+A G G LPGP P PR AG GPP  
Sbjct: 704 GPPGSAGSPGKDGGLNGLPGPIG--PPGPRGRTGDAGAPGP 742

Score = 90 (13.5 bits), Expect = 2.8e-01, P = 2.5e-01  
Identities = 45/134 (33%), Positives = 53/134 (39%)

Query: 7 GEAGGPGAAWARRAAALPGTAAGPPRPAAPPGAAPARGGPAPGAPAQALPRSQRGRO-LA 65  
G G PG A + A G A P P P G A RG G P Q R +RG L  
Sbjct: 485 GPPGPPGPAKEGAPGADGAPAGPTGP-PQGIAGQRG--VVGLPQQ--RGERGFPLP 538

Query: 66 ERNGRPRRH--RGALAQPGHPGDLA---AGV---GR-GAGGGHRRGRHHHVRSLADL 114  
+G P + GA + G PG + AG GR GA G GR + D  
Sbjct: 539 GPSGEPGKQPSGASGERGPPGPMGPPGLAGPPGESGREGAPGAEGSPGRDGSAGKGR 598

Query: 115 LQL-PGAAGAGDRGHLPGP 133  
+ P A G PGP  
Sbjct: 599 GETGPAGAPGPPGAPGAPGP 618

Score = 83 (12.5 bits), Expect = 1.8e+00, P = 8.3e-01  
Identities = 49/156 (31%), Positives = 56/156 (35%)

Query: 7 GEAGGPGAAWARRAAALPGTAA--GPPRPAAPPGAAPARG--GPAP--GAPAQALPRSQR 60  
G+AG GA A + G GPP PA PG G GPA GAP R +  
Sbjct: 311 CDAGPKGADGAPGKDGVRGLTGPIGPPGAPAGPDGKEAGPSGPACTRGAPGD---RGE 367

Query: 61 GRQLAERNRPRRRHREGALAQPGHPGDLAAGVGRGAGGGHRRGRHHHVRSLADLLQLPGA 120  
P G G PGD A G G G + ++ PG  
Sbjct: 368 GPPGPAGFAGPPGADGQPGAKGEPGDAGAKGDAGPPGAPGPPGPIGNVG---APGP 423

Query: 121 AEGAGDRGHLPGPDARDPELPRVFLP----LAGLRGPPAAVRE 160  
G G PG RV P AG GPP A +E  
Sbjct: 424 KGARGSAGP-PGATGFPGAAGRVGPPGPGSNAGPPGPPGPAGKE 466

Score = 82 (12.3 bits), Expect = 2.3e+00, P = 9.0e-01  
Identities = 46/148 (31%), Positives = 52/148 (35%)

Query: 7 GEAGGPGAAWARRAAALPGTAAGPPRPAAPPGAAPARGGPAPGAPAQALPRSQRGRLAE 66  
G+AG PGA ++ A L G G A PG RG P A P R L  
Sbjct: 275 GDAGAPGAPGSQAGPLQGMP-GERGAAGLPGKDRGDAGPKG-ADGAPGKDGVRGLTG 332

Query: 67 RNRGRPRRRHREGALAQPGHPGDLAAGVGRGAGGGHRRGRHHHVRSLADLLQLPGAAGAGD 126  
G P G PG G+ G G RG A PGA G  
Sbjct: 333 PIGPP---GPAGAPGDKGAGPSGPACTRGAPGDRGEPGPPGP-AGFAGPPGADGQPGA 387

Query: 127 RGHLPGP-DARDPELPRVFLPLAGLRGPP 154  
+G PG A+ P P AG GPP  
Sbjct: 388 KGE-PGDAGAKGDAGPPG--P-AGPAGPP 412

Peptide information for frame 3

-----

ORF from 12 bp to 755 bp; peptide length: 248  
 Category: similarity to known protein  
 Classification: unset  
 Prosite motifs: LEUCINE\_ZIPPER (17-39)  
 LEUCINE\_ZIPPER (24-46)

## BLASTP hits

No BLASTP hits available

## Alert BLASTP hits for DKFZphtes3\_18f3, frame 3

TREMBL:AF070675.1 product: "TNF-inducible protein CG12-1"; Homo sapiens TNF-inducible protein CG12-1 mRNA, complete cds., N = 1, Score = 135, P = 1e-06

TREMBL:HS6802.1 gene: "dJ6802.1"; product: "dJ6802.1"; Homo sapiens DNA sequence from PAC 6802 on chromosome 22. Contains apolipoprotein L, myosin heavy chain, ESTs, CA repeat, STS and GSS., N = 1, Score = 107, P = 0.0023

>TREMBL:AF070675.1 product: "TNF-inducible protein CG12-1"; Homo sapiens TNF-inducible protein CG12-1 mRNA, complete cds.  
 Length = 331

## HSPs:

Score = 135 (20.3 bits), Expect = 1.0e-06, P = 1.0e-06  
 Identities = 30/103 (29%), Positives = 55/103 (53%)

Query: 30 RLHRQVLRRLREVARRLRLRRSLVANVAGSSLSATGALAAIVGLSLSPVTLGTSLVSA 89  
 ++ + +LR +A +E + R ++NV SS A + ++ GL L+P T GTSL ++A  
 Sbjct: 91 KIQESIEKLRALANGIEEVHRGCTISNVSSSTGAASGIMSLAGLVLPFTAGTSLALTA 150

Query: 90 VGLGVATAGGAVTITSDL-SLIFCNSRELRRVQEIATCQDOMR 132  
 G+G+ A IT+ + + +S E + AT D+++  
 Sbjct: 151 AGVGLGAASAVTGITTSIVENSYTSSAEAE-ASRLTATSIDRLK 193

## Pedant information for DKFZphtes3\_18f3, frame 2

## Report for DKFZphtes3\_18f3.2

{LENGTH} 193  
 {MW} 19708.24  
 {pI} 11.90  
 {KW} All\_Alpha  
 {KW} LOW\_COMPLEXITY 55.44 %

SEQ TEVNGNGEAGGPGAARAAALPGTAAGPPRPAAPGAAPARGGPAPGAPAAQALPRSOR  
 SEG .....XXX...  
 PRD cccccccccccccchhhhhhhccccccccccccccccccccccccccccchhhhh

SEQ GRQLAERNRPRRRHGALAQPGHPGDLAGVGRGAGGGHSRRGRHHVRSADLLQLPGA  
 SEG .....XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX.....  
 PRD hhhhhccchhhhhhhcccccc

SEQ AEGAGDRGHLPGPDARDPELPRVFLPLAGLRGPPAAAVREERLHRPVQFCLLHRLMLTW  
 SEG .....XXXXXXXXXXXXX.....XXXXXXXXXXXXX  
 PRD cccccccccccccccccccccccccccccccccccccchhhhhhhccccchhhhhhhhhhhc

SEQ LPHPOAGGGGHQ  
 SEG xxxxxxxxxxxxx  
 PRD ccccccccccc

(No Prosite data available for DKFZphtes3\_18f3.2)

(No Pfam data available for DKFZphtes3\_18f3.2)

## Pedant information for DKFZphtes3\_18f3, frame 3

[illegible]

PS00029	17->39	LEUCINE_ZIPPER	PDOC00029
PS00029	24->46	LEUCINE_ZIPPER	PDOC00029

(No Pfam data available for DKFZphtes3\_18f3.3)

DKFZphtes3\_1817

-----

group: cell structure and motility

DKFZphtes3\_1817 encodes a novel 1050 amino acid protein with weak partial similarity to ankyrins.

The novel protein contains an ATP/GTP-binding site motif A (P-loop) and an Ank repeat. Ankyrins are peripheral membrane proteins which interconnect integral proteins with the spectrin-based membrane skeleton. Thus the novel protein seems to be involved in coupling of cyto skeleton and cell membrane.

The new protein can find application in modulation of cyto skeleton-membrane interactions.

similarity to ankyrins

Sequenced by MediGenomix

Locus: unknown

Insert length: 4501 bp

Poly A stretch at pos. 4423, no polyadenylation signal found

```

1  GATCGCCGCG CGAGGGTGGT GGGCATCGAG GTCCAGCAG CGGACGAGGG
51  AGGTGCCGCG GTGCCGCGG ATGGGCTGGG AATGAAGCGA TGTAAGCTTT
101 TAAGAGATTT GCTCTGACCC ATCTGAAGTC CATATGGCTC TGATGATGA
151 AGACCTCCCTG AAAAATCCTT TCTATCTGGC TCTGAAAAG TCGCGCCCTG
201 ACTTGTGCGAG CAAAGTGCGC CAAATCCATG GCATTGTCTT AGTACCCCTGC
251 AAGGAAGCC TCTCGAGCAG CATCCAGTCT ACTTGTCACT TTGAGTCTTA
301 CATTITGATA CCTCTGGAAG AGCATTTTCA GACCTTAAAT GGAAAGGATG
351 TCTTTATTCA AGGGAACAGG ATTAATTAAG GAGCTGGTTT TGCTGTCTT
401 CTCTCAGTGC CCATTCTCTT TGAAGAACT TTCTACAATG AAAAAGAAGA
451 GAGTTTCAGC ATCCTGTGTA TAGCCATCC TTTGGAAAAG AGAGAGAGTT
501 CAGAAGAGCC TTTGGCACCC TCAGATCCCT TTTCCCTGAA AACCATTGAA
551 GATGTGAGAG AGTTCTTGGG AAGACACTCC GAGCGATTGG ACAGGAACAT
601 CGCCTCTTTC CATCGAACAT TCCGAGAATG CGAGAGAAAG AGCCTCCGTC
651 ACCACATAGA CTCAGCGAAT GCTCTCTACA CCAATGCCT CCAGCAGCTT
701 CTGAGGAGCT CTCACCTGAA AATGCTCGCC AAGCAGGAGG CCCAGATGAA
751 CCTGATGAAG CAGGCACTGG AGATATACGT CCATCATGAA ATTTACAACC
801 TGATCTTTAA ATACGTGGGG ACCATGGAGG CAAGTGAAGA TGGGCGCTTT
851 AACAAAATCA CAAGAAGCCT TCAAGATCTT CAGCAGAAAG ATATTGGTGT
901 GAAACCGGAG TTCAGCTTTA ACATACCTCG TGCCAAAAGA GAGCTGGCTC
951 AGCTGAACAA ATGCACCTCC CCACAGCAGA AGCTTGTCTG CTTCGCAAAA
1001 GTGGTGACGC TCATTACACA GTCTCAAGC CAGAGAGTGA ACCTGGAGAC
1051 CATGTGTGCT GATGATCTGC TATCAGTCTT GTTATACCTG CTGTGAAAAA
1101 CGGAGATCCC TAATTGGATG GCAAAATTGA GTTACATCAA AAACCTCAGG
1151 TTTAGCAGCT TGGCAAAGGA TGAAGTGGGA TACTGCTGTA CCTCATTCGA
1201 AGCTGCCATT GAATATATTC GGCAAGGAAG CCTCTCTGCT AAACCCCTG
1251 AGCTGAGGGG ATTTGGAGAC AGGCTGTGCC TTAAGCAGAG AATGAGCTTA
1301 CTCTCTCAGA TGACTTCGTC TCCACCGGAC TGCTGTGTTA AGCACATTGC
1351 ATCAGGTAA CAGAAAGAAG TGGAGAGACT TCTGAGCCAA GAGGACCATG
1401 ATAAAGATAC CGTCCAAAAG ATGTGTCAAC CTCTCTGCTT CTGCGATGAC
1451 TGTGAGAAAC TCGTCTCTGG GAGGTTGAAT GATCCCTCAG TTGTCACTCC
1501 ATTTCTCAGA GACGACAGGG GGCACACCCC TCTCCATGTG GCTGCTGTCT
1551 GTGGGCGAGG ATCCCTCATC GACCTCCTGG TTTCCAAAGG CGCATGGTA
1601 AATGCCACAG ACTACCATGG GGCCACTCCG CTCCACCTGG CCTGTAGAA
1651 GGGTACCAG AGCGTGACGC TGCTGTGCT GCACCTAAG GCCAGCGCGG
1701 AAGTGCAGGA CAACAATGGG AATACGCCAC TCCACTGGC CTGCACCTAC
1751 GGCCACGAGG ACTGTGTGAA GGCTCTGGTT TACTACGAGC TGGAGTCGTG
1801 CAGACTTGAC ATTGGCAATG AGAAAGGAGA CACCCCTCTA CACATTGCTG
1851 CCCGCTGGGG CTACCAAGGC GTCATAGAGA CATTGTGCGA GAACGGAGCG
1901 TCCACCGAGA TCCAGAACAG ACTGAAGGAG ACGCCCTCTA AGTGTGCATT
1951 AAACCTCAAG ATTCTGTCTG TAATGGAAGC CTATCAGCTG TCCTTCGAGA
2001 GGAGGCAGAA GTCGTCCGAG GCCCTGTGCG AGTCCCGGCA GCGCTCCGTG
2051 GACTCCATCA GCCAAGATGC CTCCACTTCC AGCTTCTCTT CCATGTCAGC
2101 CGGCTCAAGG CAGGAGGAGA CCAAGAAGGA CTACAGAGAG GTAGAAAAAC
2151 TTTTGAGAGC AGTTGTCTGAT GGAGATCTAG AAATGGTGCG TTACTCTGTT
2201 GAATGGACAG AGGAGGACCT GGAGGATGCG GAGGACACTG TCAGTGCAGC
2251 AGACCCCGAA TTCTGTCAAC CGTTGTGCCA GTGCCCCAAG TGTGCCCCAG
2301 CTCAGAAGAG GCTGGCGAAG GTTCTTGCCA GTGGGCTTGG TGTGAACGTG
2351 ACCAGCCAGG ACGGCTCTCT CCCGCTGCACT GTGCGCGCCC TGACGCGCGG
2401 GCGGACCTC ATCCGCTCTC TGCTGAAGCA CGGGGCCAAC GCAGGTGCCA
2451 GGAACGCAGA CCAAGCCGTC CCGCTCCACC TGGCTGCCA GCAGGGCCAC
2501 TTTCAAGTGG TGAAGTGTCT GTTAGATTCT AATGCAAAAC CCAATAAGAA
2551 GGACCTCAGT GGAACACAGC CCCTCATTTA CGCCTGCTCC GGTGGCCATC
2601 ACGAGCTTGT GGCACTGTCT CTACAGCAGC GGGCTTCCAT TAACGCTTCT
2651 AACAATAAGG GCAACACAGC GCTGCACAGG GCTGTGATTG AAAAGCACGT

```



```
2701 CTTCTGGTGA GAGCTGCTTC TGCTCCACGG AGCGTCAGTT CAGGTGCTGA
2751 ACAAGCGGCA GCGCACGGCT GTAGACTGTG CTGAACAGAA TTCAAAAATA
2801 ATGGAATTGC TTCAGGTGGT ACCAAGCTGT GTTGCTTCAT TAGATGATGT
2851 GGCTGAAACT GACCGCAAGG AGTATGTAC TGTAAAGATC AGGAAAAAAT
2901 GGAACCTCAA ACTGTATGAT CTACCAGATG AGCCTTTTAC AAGACAGTTT
2951 TACTTTGTCC ACTCAGCTGG TCAGTTTAAG GGAAAGACTT CAAGGGAGAT
3001 TATGGCAAGA GATAGAAGTG TCCCTAATTT AACCGAAGTT TCTTTGCATG
3051 AGCCAGGGAG GCAAGTGTG ACACGTGAGC AGAATAACCT GCCAGCTCAG
3101 AGTGGATCTC ATGCTGCTGA GAAAGGCAAC AGCGACTGGC CAGAGAGGCC
3151 TGGACTGACA CAGACTGGCC CTGGACACAG ACAGGATGCTG CGGAGACACA
3201 CGGTAGAGGA TCGGTCGTG TCCAGGGGCC CGGAGGCTGC TGGCCCCCTC
3251 TCCACTCCCC AAGAGGTTAG TGCTCCCCGG TCCTAACAGG AATGAGGAGT
3301 TGTGTAACCC ACTGCTAGGA AGCAAGGATG CAACAAGATG ATGCTGAGCG
3351 TGAACACATC TGAGAACTAA ATGTGCTTCC ATGAGACTGG CTTGAGAAGT
3401 CTTTCAGCAC AAGTTCCTGA AAGCTTTTCT GTGGCAGGAA AGAATGCAAC
3451 AAAAAAGTTA ACCACCACCA TCTCTCTCCT CTTCAAAGCT AATGAATACA
3501 ATTGAAACAG ACAAATAATC CAGTAGCATC CAGATCCTTA AGCCAGAGGT
3551 GCATGCTTCT TTTTAAGTAT GAGGGTTTGT TGGTCACAGT GGGAGAGGTT
3601 TCACCACCGC ATTCTGACCT CCTCCTCCCA AAAGGTGCTA AACCTCTCTG
3651 ACCTGTGTAC ATTCAACAAC CACAGCTAGA ATTCCTCCAC CTAGGATTAA
3701 GCTGGAGAGA AGTAAGTAAT TTAGGTTTCA TGGTACTGTA GAGGCCAGGC
3751 TGAATGTCA TATCTGAAGG AAGAAAGCAG CAGCTGGACA ATGTTTCTTT
3801 GCAAAGCAAC ACTCGAACCA AAAGATGCCT CAATCCCATT TTGATATTCA
3851 TTTTAGTGAA AGGATGCATC AGACCTGTTC CACATCATGC ACATGGGAAA
3901 GGGTGGTTAT CATTTTCCTT CTAACAAGTA GGTACAGATA TTCGGTTACT
3951 ACACGTGCAC CTGTAGCAGT ATTTCTAGAA ACATCCCCTT TTGTTGAGAA
4001 CCTCCCTTGA ATGTCTGTCA CACTCACACC TGACGGGATG GTTACTGGAT
4051 TAGAGAGTAG ATTTGGCACA TCTTTTCTTA GTCTTTTGAT TCAAATTCRA
4101 AACTTAACAG CACAAACCAG GTCAGAGTTA CTTTCGGTTA GAATTTATTG
4151 CCATTTATTCT CTTTTTATAA ATTTCTATAG ATTATACGTG TATTTTATTG
4201 TTATTGGCCT AGAGCTACAC GTATATGGGT TTGTCTCTAG TCCGTTTTCA
4251 AATGACCTTG TGATAGGGAA ATGGTTTTGT CCATGTTCTT GGAAATACTT
4301 GTGTATGTAC AGAAGGAAGG GAGGGATTAT TTTCTACAA AGTAATTAT
4351 GATTCTTAAT TTTCTAATGT GCCTTGGATA TGTGCCAAAT GATGGAAAAG
4401 AAACAGTAAA CTTTATGATT CTTAAAAAAA AAAAAAATAA AAAAAAATAA
4451 AAAAAAATAA AAAAAAATAA AAAAAAATAA AAAAAAATAA AAAAAAATAA
4501 G
```

## BLAST Results

No BLAST result

## Medline entries

No Medline entry

## Peptide information for frame 2

ORF from 134 bp to 3283 bp; peptide length: 1050  
Category: similarity to known protein  
Classification: Cell structure/motility  
Prosite motifs: ATP\_GTP\_A (945-953)

```
1 MALYDEDLLK NPFYLAQKC RPDLCSEKVAQ IHGIVLVPCX GSLSSSIQST
51 CQFESYILIP VEEHFQTLNG KDVFIQGNRI KLGAGFACLL SVPILEETTF
101 YNEKEESFSI LCIAHPLEKR ESSEELAPS DPFLSKTIED VREFLGRHSE
151 RFDNRNIAFSH RTFRECEKRS LRHHIDSANA LYTKCLQQLL RDSHLKMLAK
201 QEAQNLNMQQ AVEIYVHHEI YNLIKYVGT MEASEDAAFN KITRSLQDLQ
251 QKDIGVKEPF SFNIPRAKRE LAQLNKCTSP QOKLVCLRKV VOLITQSPSQ
301 RVNLETMCAD DLLSVLLYLL VKTEIPNWM A NLSYIKNFRF SSLAKDELGY
351 CLTSFEAAIE YIRQGSLSAK PPESEGFQDR LFLKQRMILL SMTSSPTDC
401 LFKHIASGNQ KEVERLLSQE DHDKDTVQKM CHPLCFDCCD EKLVSGRIND
451 PSVVTPTFSRD DRGHTPLHVA AVCGQASLID LLVSKGAMVN ATDYHGATPL
501 HLACQKGYQS VTLLLLHYKA SAEVDNNGN TPLHLACTYG HEDCVKALVY
551 YDVESCRLDI GNEKGDTPH IAAEWGQGV IETLLQNGAS TEIQNRKLET
601 PLKCALNSKI LSVMEAYHLS FERRQKSEA PVQSPQRSVD SISQESSTSS
651 FSSMSAGSRQ EETKKDYREV EKLLRAVADG DLEMVRYLLE WTEEDLEDAE
701 DTVSADPEF CHPLCQCPKC APAQKRLAKV PASGLGVNVT SQDGSPLPHV
751 AALHGRADLI RLLKKGANA GARNADQAVP LHLACQGHF QVVKCLDSN
801 AKPNKKDLGS NTPLIYACSG GHHELVALLL QHGASINASN NKGNTALHEA
851 VIEKHVFPVE LLLHGAASVQ VLNKRORTAV DCAEQNSKIM ELLQVVPSCV
```

901 ASLDDVAETD RKEYVTVKIR KWNLSKLYDL PDEPFTROFY FVHSAGQFKG  
951 KTSREIMARD RSVPNLTEGS LHEPGRQSVT LRQNNLPAQS GSHAAEKGNS  
1001 DWPERPCLTQ TGPCHRRMLR RHTVEDAVVS QGPEAAGPLS TPQEVSAERS

## BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKF2phtes3\_1817, frame 2

TREMBL:HSU43965.1 gene: "ANK3"; product: "ankyrin G119"; Human ankyrin  
G119 (ANK3) mRNA, complete cds., N = 2, Score = 287, P = 3.7e-21

PIR:I49502 ankyrin - mouse, N = 3, Score = 365, P = 2.2e-27

TREMBL:HSANKY\_2 product: "alt. ankyrin (variant 2.2)"; Human mRNA for  
ankyrin (variant 2.1), N = 2, Score = 380, P = 7.3e-31

SWISSPROT:ANK1\_HUMAN ANKYRIN R (ANKYRINS 2.1 AND 2.2) (ERYTHROCYTE  
ANKYRIN), N = 2, Score = 380, P = 8.2e-31

PIR:SJHUK ankyrin 1, erythrocyte splice form 1 - human, N = 2, Score =  
380, P = 8.2e-31

>TREMBL:HSANKY\_2 product: "alt. ankyrin (variant 2.2)"; Human mRNA for  
ankyrin (variant 2.1)  
Length = 1,719

## HSPs:

Score = 380 (57.0 bits), Expect = 7.3e-31, Sum P(2) = 7.3e-31  
Identities = 139/447 (31%), Positives = 207/447 (46%)

Query: 462 RGHTPLHVAAVCGQASLIDLLVSKGAMVNATDYHGATPLHLACQKGYQSVTL LLLHYKAS 521  
+G+T LH+AA+ GQ ++ LV+ GA VNA G TPL+ A Q+ + V LL A+  
Sbjct: 77 KGNTALHIALAGQDEVVRELNVYGANVNAQSQKGTPLYMAAQENHLEVVKFLLENGAN 136

Query: 522 AEVDNNGNTPLHLACTYGHEDCVKALVYDVES-CRL----- 558  
V +G TPL +A GHE+ V L+ Y + RL  
Sbjct: 137 QNVATEDGFTPLAVALQQGHENVVVAHLINYGTKGKVRPLALHIAARNDDTRTAAVLLQND 196

Query: 559 ---DIGNEKGDTPHLHIAARWYQGVETLLQNGASTEIQNRLKETPLKCALNSKILSVME 615  
D+ ++ G TPLHIAA + V + LL GAS + TPL A S+ +V+  
Sbjct: 197 PNPVLSKTGFTPLHIAAHYENLNVAQLLNKGASVNFPTONGITPLHIA--SRAGNVIM 254

Query: 616 AYHLSFERRQKSSEAPVQSPQRSVDSISQESSTS-SFSSMSAGSR-QEETKKDYREVEKL 673  
L +R + E + + ++ S + G+ Q +TK +  
Sbjct: 255 V-RLLDGRGAQI-ETTKDELTPHCAARNGHVRRISEILLDHGAPIQAKTKNGLSPIHM- 311

Query: 674 LRAVADGD-LEMVRYLLEWTEEDLEDAEDTVSAADPEFCHPLCQCPKCAPAQRRLAKVPA 732  
A GD L+ VR LL++ E ++D T+ P H C R+AKV  
Sbjct: 312 ---AAQGDHLDVRLLLQYDAE-IDDI--TLDHLP--LHVAHC-----GHHRVAKVLL 358

Query: 733 S-GLGVNVTSDQSSPLHVAALHGRADLIRLLKHGANAGARNADQAVPLHLACQGHFQ 791  
G N + +G +PLH+A ++ LLLK GA+ A PLH+A GH  
Sbjct: 359 DKGAKPNSRALNGFTPLHIACKKNHVRVMEILLKTGASIDAVTESGLTPLHVASFHGLP 418

Query: 792 VVKCLLDSNAKPNKKOLSGNTPLIYACSGGHHELVALLLQHGASINASNKNGTALHEAV 851  
+VK LL A FN +++ TPL A GH E+ LLQ+ A +NA T LH A  
Sbjct: 419 IVNLLQRGASPNVSNVKVET-LHMAARAGHTEVAKYLLQNKAKVNAKAKDDQTPHCAA 478

Query: 852 IEKHVFVVELLLHGASVQVNLKRQRTAVDCAEQNSKIMELLQVV 896  
H +V+LL + A+ + T + A + + +L ++  
Sbjct: 479 RIGHTNMVKLLLENNANPNLATTAGHTPLHIAAREGHVETVLALL 523

Score = 378 (56.7 bits), Expect = 1.2e-30, Sum P(2) = 1.2e-30  
Identities = 130/447 (29%), Positives = 195/447 (43%)

Query: 465 TPLHVAAVCCQASLIDLLVSKGAMVNATDYHGATPLHLACQKGYQSVTL LLLHYKASAEV 524  
TPLH AA G + ++L+ GA + A +G +P+H+A Q + LLL Y A +  
Sbjct: 274 TPLHCAARNGHVRRISEILLDHGAPIQAKTKNGLSPIHMAAQGDHLDVRLLLQYDAEIDD 333

Query: 525 QDNNGNTPLHLACTYGHEDCVKALVYDVE-----SCR----- 557  
+ TPLH+A GH K L+ + +C+  
Sbjct: 334 ITLDHLPPLHVAACHGHRVAKVLLDKGAKPNSRALNGFTPLHIACKKNHVRVMEILLKT 393

Query: 558 ---LDIGNEKGDTPHLHIAARWYQGVETLLQNGASTEIQNRLKETPLKCALNSKILSV 614  
+D E G TPLH+A+ G+ +++ LLQ GAS + N ETPL A + V  
Sbjct: 394 GASIDAVTESGLTPLHVASFHGLPIVKNLLQRGASPNVSNVKVETPLHMAARAGHTEVA 453

Query: 615 EAYHLSFERRQKSSEAPVQSPQRSVDSISQESSTSSFSMSAGSRQEETKKDYREVEKLL 674  
 + Y L + + + Q+P I + +A T L  
 Sbjct: 454 K-YLLQNKAKVNAKAKDDQTPHCAARIGHTNMVKLLLENNANPNLATTAGH---TPLH 508

Query: 675 RAVADGDLEMYRYLLEWTEEDLEDAEDTVSAADPEFCHPLCQCPKCAPAQKRLAKVPASG 734  
 A +G +E V LLE ++ A T P H + K A+ L +  
 Sbjct: 509 IAREGHVETVLALLE---KEASQACMTKKGFTP--LHVAAKYGKVRVAELLER----D 559

Query: 735 LGVNVTSQDGSSPLHVAALHGRADLIRLLKKHGANAGARNADOAVPLHLACQGHFQVVK 794  
 N ++G +PLHVA H D++LLL G + + PLH+A +Q +V +  
 Sbjct: 560 AHPNAAAGKGLTPLHVAVHHNNLDIVKLLPRGGSPPHSPAWNNGYTPHIAAKQNQVEVAR 619

Query: 795 CLDLSNAKPNKKDLGNTPLIYACSGGHHHELVALLLQHGASINASNKGTALHEAVIEK 854  
 LL N + + G TPL A GH E+VALL A+ N N G T LH E  
 Sbjct: 620 SLLQYGGSSANAESVQGVTPHLAAQEGHAEMVALLSKQANGNKNKSGLTPLHLVAQEG 679

Query: 855 HVFVVELLLLHGASVQVLNKRQRTAVDCAEQ--NSKIMELL 893  
 HV V ++L+ HG V + T + A N K+++ L  
 Sbjct: 680 HVPVADVLIKGVMDATTRMGYTPHVASHYGNIKLVKFL 720

Score = 367 (55.1 bits), Expect = 1.8e-29, Sum P(2) = 1.8e-29  
 Identities = 131/489 (26%), Positives = 210/489 (42%)

Query: 404 HIAS--GNQKEVERLLSQEDHDKDTVQKMCHPL-CFDDCEKLVSGRLNDPSVVTFFSRD 460  
 HIAS GN V LL + + + PL C + +S L D ++  
 Sbjct: 244 HIASRRGNVIMVRLLLDRGAQIETKTKDELTPHCAARNGHVRISIELLDHGAPIQ-AKT 302

Query: 461 DRGHTPLHVAACVCGQASLIDLLVSKGAMVNATDYHGATPLHLACQKGYSVTLTLLHYKA 520  
 G +P+H+AA + LL+ A ++ TPLH+A G+ V +LL A  
 Sbjct: 303 KNGLSPIHMAAQGDHLDVRLLLQYDAEIDDITLOHLTPLHVAACHGHHRVAVLLDKGA 362

Query: 521 SAEVQDNNGNTPLHLACTYGHEDCVKALVYDVESCRLDIGNEKGDTPHIAARWYGQV 580  
 + NG TPLH+AC H ++ L+ +D E G TPLH+A+ G+ +  
 Sbjct: 363 KPNSRALNGFTPLHIACKKNHVRVMELLLK---TGASIDAVTESGLTPLHVASFHGLPI 419

Query: 581 IETLLQNGASTEIQNRKETPLKCAL---NSKILSVMEAYHLSFERRQKSSEAPVQSPQR 637  
 ++ LLQ GAS + N ETP L A ++++ + + K + P+ R  
 Sbjct: 420 VKNLLQRGASPNVSNVKVETPLHMAARAGHTEVAKYLLQNKAKVNAKAKDDQTPHCAAR 479

Query: 638 ---SVDSISQESSTSSFSMSAGSRQEETKKDYREVEKLLRAVADGDLEMYRYLLEWTE 693  
 ++ + E++ + +AG VE +L + + +T  
 Sbjct: 480 IGHNTNMVKLLLENNANPNLATTAGHTPLHIAAREGHVETVLALLEKEASQACMTKKGFTP 539

Query: 694 EDLEDAEDTVSAAD---PEFCHPLCQ-----CP-KCAPAQKRLAKVPA---SGLGVNVTS 741  
 + V A+ HP P A L V G + +  
 Sbjct: 540 LHVAAKYGKVRVAELLERDAHPNAAAGKNGLTPLHVAVHHNNLDIVKLLPRGGSPPHSPA 599

Query: 742 QDGSSPLHVAALHGRADLIRLLKKHGANAGARNADOAVPLHLACQGHFQVVKCLLDSNA 801  
 +G +PLH+AA + ++ R LL++G +A A + PLHLA Q+GH ++V LL A  
 Sbjct: 600 WNGYTPHIAAKQNQVEVARSLQYGGSSANAESVQGVTPHLAAQEGHAEMVALLSKQA 659

Query: 802 KPNKKDLSGNTPLIYACSGGHHHELVALLLQHGASINASNKGTALHEAVIEKHVVFVEL 861  
 N + SG TPL GH + +L++HG ++A+ G T LH A ++ +V+  
 Sbjct: 660 NGNLGNKSGLTPLHLVAQEGHVPVADVLIKGVMDATTRMGYTPHVASHYGNIKLVKF 719

Query: 862 LLLHGASVQVLNK 874  
 LL H A V K  
 Sbjct: 720 LLHQADVNATK 732

Score = 345 (51.8 bits), Expect = 4.2e-27, Sum P(2) = 4.2e-27  
 Identities = 146/506 (28%), Positives = 233/506 (46%)

Query: 404 HIAS--GNQKEVERLLSQEDHDKDTVQK---MCHPLCFDDCEKLVSGRLNDPSVVTFFS 458  
 H+AS G+ K V LL +E + T +K H +++V +N + V +  
 Sbjct: 50 HLASKEGHVMVMVELLHKEIILETTTKGNTALHIAALAGQ-DEVVRELNVYGANVN--A 106

Query: 459 RDDRGTPLHVAACVCGQASLIDLLVSKGAMVNATDYHGATPLHLACQKGYSVTLTLLHY 518  
 + +G TPL++AA ++ L+ GA N G TPL +A Q+G+++V L++Y  
 Sbjct: 107 QSQKGFTPLMAAQENHLEVVKFLENGANQNVATEDGFTPLAVALQGHENVAHLINY 166

Query: 519 KASAEVQDNNGNTP-LHLACTYGHEDCVKALVYDVESCRLDIGNEKGDTPHIAARWGY 577  
 +V+ P LH+A ++D A V + D+ ++ G TPLHIA +  
 Sbjct: 167 GTKGKVR-----LPALHIAAR--NDDRTAAVLLQNDP-NPDVLSKTGFTPLHIAAHYEN 218

Query: 578 QGVETLLQNGASTEIQNRKETPLKCAL---NSKILSVMEAYHLSFERRQKSSEAPVQS 634  
 V + LL GAS + TPL A N ++ ++ E + K P+  
 Sbjct: 219 LNVAQLLNRGASVNFPTQNGITPLHIAARRGNVIMVRLLLDRGAQIETKTKDELTPHLC 278

Query: 635 PQRSVDSISQESSTSSFSMSAGSRQEETKKDYREVEKLLRAVADGD-LEMYRYLLEWTE 693  
 R+ E + + A +TK + A GD L+ VR LL++  
 Sbjct: 279 AARNGHVRISIELLDHGAPIQA-----KTKNGLSPIHM---AAQGDHLDVRLLLQYDA 329

Query: 694 EDLEDAE-DTVSAAD-PEFC--HPLCQC-----PK-----CAPAQKRLAK 729  
E++D D++ C H++ P C R++  
Sbjct: 330 E-IDDTLDHLTPLHVAHCGHHRVAKVLLQKAKPNSRALNGFTPLHIACKKNHVRVME 388

Query: 730 VPA-SGLGVNVTSDQSSPLHVAALHGRADLIRLLKKGANAGARNADQAVPLHLACQQG 788  
+ +G ++ ++ G +PLHVA+ G +++ LL+ GA+ N PLH+A + G  
Sbjct: 389 LLLKTGASIDAVTESGLTFLHVASFMGHLPIVKNLLQRGASPNVSNVKVETPLHMAARAG 448

Query: 789 HFQVVKCLLDSNAKPNKDLGNTPLIYACSGGHHVALLLQHGASINASNKGNLTALH 848  
H +V K LL + AK N K TPL A GH +V LLL++ A+ N + G+T LH  
Sbjct: 449 HTEVAKYLLQNKAKVNAKAKDDQTPHCAARIGHTNMVKLLLENNANPNLATTAGHTPLH 508

Query: 849 EAVIEKHVFVVELLLHGASVQVLNKRQRTAVDCAEQNSKIM--ELL 893  
A E HV V LL AS + K+ T+ A+ K+ ELL  
Sbjct: 509 IAAAREGHVETVLALLEKEASQACMTKKGFTPLHVAARYGKVRVAELL 555

Score = 243 (36.5 bits), Expect = 1.6e-14, Sum P(2) = 1.6e-14  
Identities = 64/199 (32%), Positives = 97/199 (48%)

Query: 404 HIAS--GNQKEVERLLSQEDHDKDTVQKMCHPLCFDCEKLVSGRLNDPSVVTFFSRDD 461  
H+A+ G + E LL ++ H + PL L +L P +P S  
Sbjct: 541 HVAARYGKVRVAELLERDAHNAAGKNGLTPLHVAVHNNLDIVKLLPRGSGPHSPAW 600

Query: 462 RGHTPLHVAACVQASLIDLLVSKGAMVNATDYHGATPLHLACQKGYQSVTLTLLHYKAS 521  
G+TPLH+AA Q + L+ G NA G TPLHLA Q+G+ + LLL +A+  
Sbjct: 601 NGYTPLHIAAKQNVQVARSLLQYGGSSANAESVQGVTPHLHAAQEGHAEMVALLSKQAN 660

Query: 522 AEVDNNGNTPLHLACTYGHEDCVKALVYVDVESCRLDIGNEKGDTPHLIAARWGYQGV 581  
+ + +G TPLHL GH L+ + V +D G TPLH+A+ +G ++  
Sbjct: 661 GNLGNKSGLTPLHLVAQEGHVPVADVLIKHGV--MVDATTRMGYTPLHVASHYGNIKLV 717

Query: 582 ETLLQNGASTEIQNRKLTPL 602  
+ LLQ+ A + +L +PL  
Sbjct: 718 KFLQHQADVNNAKTLGYSPL 738

Score = 242 (36.3 bits), Expect = 5.0e-29, Sum P(2) = 5.0e-29  
Identities = 63/176 (35%), Positives = 92/176 (52%)

Query: 734 GLGVNVTSDQSSPLHVAALHGRADLIRLLKKGANAGARNADQAVPLHLACQQGHFQVV 793  
G VN T Q+G +PLH+A+ G ++RLLL GA + D+ PLH A + GH ++  
Sbjct: 229 GASVNFTPQNGITPLHIASRGNVIMVRLLDLQGAQIETKTKDELTPHCAARNGHVRIS 288

Query: 794 KCLLDSNAKPNKDLGNTPLIYACSGGHHVALLLQHGASINASNKGNLTALHEAVIE 853  
+ LLD A K +G +P+ A G H + V LLLQ+ A I+ T LH A  
Sbjct: 289 EILLDHGAPIQAKTKNGLSPIHMAAQGDHLDCVRLLLQYDAEIDDTLDHLTPLHVAHC 348

Query: 854 KHVFVVELLLHGA--SVQVLNKRQRTAVDCAEQNSKIMELLQVVPSCVASLDDVAET 909  
H V ++LL GA + + LN + C + + +MELL AS+D V E+  
Sbjct: 349 GHHRVAKVLLDKGAKPNSRALNGFTPLHIACKKNHVRVMEMLLKTG---ASIDAVTES 403

Score = 242 (36.3 bits), Expect = 3.3e-14, Sum P(2) = 3.3e-14  
Identities = 80/284 (28%), Positives = 129/284 (45%)

Query: 404 HIAS--GNQKEVERLLSQEDHDKDTVQKMCHPLCFDCEKLVSGRLNDPSVVTFFSRDD 461  
HIA+ G+ + V LL +E +K PL K+ L P +  
Sbjct: 508 HIAAREGHVETVLALLEKEASQACMTKKGFTPLHVAARYGKVRVAELLERDAHNAAGK 567

Query: 462 RGHTPLHVAACVQASLIDLLVSKGAMVNATDYHGATPLHLACQKGYQSVTLTLLHYKAS 521  
G TPLHVA ++ LL+ +G ++ ++G TPLH+A ++ V LL Y S  
Sbjct: 568 NGTTPHVAVHNNLDIVKLLPRGSGPHSPAWNGLYTPHIAAKQNVQVARSLLQYGS 627

Query: 522 AEVDNNGNTPLHLACTYGHEDCVKALVYVDVESCRLDIGNEKGDTPHLIAARWGYQGV 581  
A + G TPLHLA GH + V L+ ++GN+ G TPLH+ A+ G+ V  
Sbjct: 628 ANAESVQGVTPHLHAAQEGHAEMVALLSKQANG--NLGNKSGLTPLHLVAQEGHVPVA 684

Query: 582 ETLLQNGASTEIQNRKLTPLKCAL---NSKILSVMEAYHLSFERRQKSSEAPV-QSPQR 637  
+ L++G + R+ TPL A N K++ + + + K +P+ Q+ Q+  
Sbjct: 685 DVLIKHGMVDATTRMGYTPLHVASHYGNIKLVKFLQHQADVNNAKTLGYSPLHQAQQ 744

Query: 638 S-VDSISQ--ESSTSFSSMSAGSRQEETK--DYREVEKLLRAVD 679  
D ++ ++ S S G+ K Y V +L+ V D  
Sbjct: 745 GHTDIVTLLKNGASPNEVSSDGTTPLAIAKRLGYSVTDVLKVVD 791

Score = 235 (35.3 bits), Expect = 7.9e-34, Sum P(2) = 7.9e-34  
Identities = 58/165 (35%), Positives = 83/165 (50%)

Query: 734 GLGVNVTSDQSSPLHVAALHGRADLIRLLKKGANAGARNADQAVPLHLACQQGHFQVV 793  
G N S G +PLH+AA G A+++ LLL AN N PLHL Q+GH V  
Sbjct: 625 GGSANAESVQGVTPHLHAAQEGHAEMVALLSKQANGNLGNKSGLTPLHLVAQEGHVPVA 684

Query: 794 KCLDSNAKPNKKDLSGNTPLIYACSGGHHVALLLQHGASINASNKGNLTALHEAVIE 853  
L+ + G TPL A G+ +LV LLQH A +NA G + LH+A +  
Sbjct: 685 DVLIKHGVMVDATTRMGYTPLHVASHYGNIKLVKFLQHQADVNAKTLGYSPLHQAAQ 744

Query: 854 KHVFVVELLLHGGASVQVLNKRQRTAVDCAEQNS--KIMELLQV 896  
H +V LLL +GAS ++ T + A++ + ++L+VV  
Sbjct: 745 GHTDIVTLLKNGASPNEVSSDGTTPLAIAKRLGYISVTDVLKVV 789

Score = 233 (35.0 bits), Expect = 7.9e-34, Sum P(2) = 7.9e-34  
Identities = 67/202 (33%), Positives = 100/202 (49%)

Query: 404 HIAS-GNQKEVERLLSQEDHDKDTVQKMCH--PLCFDDC-EKLVSGRLNDPSVVTFFSR 459  
H+A+ G+ + RLL Q D + D + + H PL C V+ L D P SR  
Sbjct: 310 HMAAQGDHLDVRLLLQYDAEIDDIT-LDHLTPLHVAACHGHRVAKVLLDKGA-KPNSR 367

Query: 460 DDRGHTPLHVAACVQASLIDLLVSKGAMVNATDYHGATPLHLACQKGYQSVTLLLLHYK 519  
G TPLH+A +++LL+ GA ++A G TPLH+A G+ + LL  
Sbjct: 368 ALNGFTPLHIACKKNHVRVMELELLKTGASIDAVTESGLTPLHVASFMGHLPIVKNLLQRG 427

Query: 520 ASAEVQDNNNGNTPLHIACTYGHEDCVKALVYDVESCRLDIGNEKGDTPHIAARWGYQG 579  
AS V + TPLH+A GH + K L+ +++ + TPLH AAR G+  
Sbjct: 428 ASPVSNVNVETPLHMAARAGHTVAKYLLQ---NKAKVNAKAKDDOTPLHCAARIGHTN 484

Query: 580 VIETLLQNGASTEIQNRKLTPLKCA 605  
+++ LL+N A+ + TPL A  
Sbjct: 485 MVKLLLENNANPNLATTAGHTPLHIA 510

Score = 226 (33.9 bits), Expect = 7.0e-33, Sum P(2) = 7.0e-33  
Identities = 53/153 (34%), Positives = 83/153 (54%)

Query: 743 DGSSPLHVAALHGRADLIRLLKKGANAGARNADQAVPLHLACQOQGHFQVVKLLDSNAK 802  
+G +PLH+AA + ++ R LL++G +A A + PLHLA Q+GH ++V LL A  
Sbjct: 601 NYGTPHIAAKQNVQEVARSLLQYGSANAEVQGVTPHLHAAQEGHAEMVALLSKQAN 660

Query: 803 PNKKDLSGNTPLIYACSGGHHVALLLQHGASINASNKGNLTALHEAVIEKHVFVVELL 862  
N + SG TPL GH + +L++HG ++A+ G T LH A ++ +V+ L  
Sbjct: 661 GNLGNKSLTPLHLVAQEGHVPVADVLIKHGVMVDATTRMGYTPLHVASHYGNIKLVKFL 720

Query: 863 LLHGASVQVLNKRQRTAVDCAEQ--NSKIMELL 893  
L H A V K + + A Q ++ I+ LL  
Sbjct: 721 LQHQADVNAKTLGYSPLHQAAQGHDTIVTLL 753

Score = 198 (29.7 bits), Expect = 2.5e-11, Sum P(2) = 2.5e-11  
Identities = 51/157 (32%), Positives = 82/157 (52%)

Query: 737 VNVTSQDGSSPLHVAALHGRADLIRLLKKGANAGARNADQAVPLHLACQOQGHFQVVKCL 796  
+ T++ G++ LH+AAAL G+ ++R L+ +GAN A++ PL+AA Q+ H +VVK L  
Sbjct: 71 LETTTKKGNTALHIAAALQAGDEVVRELNVYGANVNAQSQKGTPLYMAAQENHLEVVKFL 130

Query: 797 LDSNAKPNKKDLSGNTPLIYACSGGHHVALLLQHGASINASNKGNLTALHEAVIEKHV 856  
L++ A N G TPL A GH +VA L+ +G ALH A  
Sbjct: 131 LENGANQWATEDGFTPLAVALQOQGHENVVAHLINYGTK---GKVRPLALHIAARNDDT 186

Query: 857 FVVELLLHGGASVQVLNKRQRTAVDCAE--QNSKIMELL 893  
+LL + + VL+K T + A +N + +LL  
Sbjct: 187 RTAAVLLQNDPNPDVLSKTGFTPLHIAAHYENLNVAQLL 225

Score = 186 (27.9 bits), Expect = 6.6e-29, Sum P(2) = 6.6e-29  
Identities = 55/143 (38%), Positives = 68/143 (47%)

Query: 463 GHTPLHVAACVQASLIDLLVSKGAMVNATDYHGATPLHLACQKGYQSVTLLLLHYKASA 522  
GHTPLH+AA G + L+ K A G TPLH+A + G V LLL A  
Sbjct: 503 GHTPLHIAAREGHVETVIALLEKEASQACMTKKGFTPLHVAAYGKVRVAELLERDAHP 562

Query: 523 EVQDNNNGNTPLHLACTYGHEDCVKALVYDVESCRLDIGNEKGDTPHIAARWGYQGVIE 582  
NG TPLH+A + + D VK L+ S N G TPLHIAA+ V  
Sbjct: 563 NAAGKNGLTPLHVAVHHNNLDIVKLLPRG-GSPHSPAWN--GYTPLHIAAKQNVQEVAR 619

Query: 583 TLLQNGASTEIQNRKLTPLKCA 605  
+LLQ G S ++ TPL A  
Sbjct: 620 SLLQYGSANAEVQGVTPHLA 642

Score = 182 (27.3 bits), Expect = 2.9e-28, Sum P(2) = 2.9e-28  
Identities = 54/185 (29%), Positives = 89/185 (48%)

Query: 738 NVTSQDGSSPLHVAALHGRADLIRLLKKGANAGARNADQAVPLHLACQOQGHFQVVKCL 797  
N+ ++ G +PLH+ A G + +L+KHG A PLH+A G+ ++VK LL  
Sbjct: 662 NLGNKSLTPLHLVAQEGHVPVADVLIKHGVMVDATTRMGYTPLHVASHYGNIKLVKFL 721

Query: 798 DSNKPNKKDLSGNTPLIYACSGGHHVALLLQHGASINASNKGNLTALHEAVIEKHVF 857  
A N K G +PL A GH ++V LLL++GAS N ++ G T L A ++

Sbjct: 722 QHQADVNAKTKLGYSPFHQAQQGHTDIVTLLKNGASPNVSSDGTTPLAIAKRLGYIS 781

Query: 858 VVELLLLHGASVQVLNKRQRTAVDCAEQNSKIMELLQVVPSCVASLDDVAETDRKEYVTV 917  
V ++L + V ++ V + S P V + DV+E + +E ++

Sbjct: 782 VTDVLKV-----VTDETSFVLVSDKHRMS-----FPETVDEILDVSEDEGEELISF 827

Query: 918 KIRKK 922  
K ++

Sbjct: 828 KAERR 832

Score = 180 (27.0 bits), Expect = 5.0e-29, Sum P(2) = 5.0e-29  
Identities = 41/121 (33%), Positives = 67/121 (55%)

Query: 486 GAMVNATDYHGATPLHLACQKGYQSVTLLLLHYKASAEVQDNNGNTPLHLACTYGHEDCV 545  
G +N + +G LHLA ++G+ + + LLH + E GNT LH+A G ++ V

Sbjct: 35 GVDINTCNQNGNLGLHLASKEGHVKMVVELLHKEIILETTTCKGNTALHIAALAGQDEVV 94

Query: 546 KALVYDVESCRLDIGNEKGDTPHIAARWGYQGVETLLQNGASTEIQNRLKETPLKCA 605  
+ LV Y ++ ++KG TPL++AA+ + V++ LL+NGA+ + TPL A

Sbjct: 95 RELVNY---GANVNAQSQKGFPLYMAAQENHLEVVKFLENGANQNVATEDGFTPLAVA 151

Query: 606 L 606  
L

Sbjct: 152 L 152

Score = 166 (24.9 bits), Expect = 3.4e-06, Sum P(2) = 3.4e-06  
Identities = 89/318 (27%), Positives = 140/318 (44%)

Query: 448 LNDPSVVPFSDRRGHTPLHVAACGQASLIDLLVSKGAMVNATDYHGATPLHLACQK 507  
L + + V ++DD+ TPLH AA G +++ LL+ A N G TPLH+A ++G

Sbjct: 457 LQNKAKVNAKAKDDQ--TPLHCAARIGHTNMVKLLLENNANPNLATTAGTPLHIAAREG 514

Query: 508 YQSVTLLLLHYKASAEVQDNNGNTPLHLACTYGHEDCVKALVYD----- 552  
+ L LL +AS G TPLH+A YG + L+ D

Sbjct: 515 HVETVLALLEKEASQACMTKKGFTPLHVAAYGKVRVAELLERDAHPNAAGNGLTPLH 574

Query: 553 --VESCRLDI-----GNE-----KGDTPHIAARWGYQGVETLLQNGASTEIQNRL 597  
V LDI G+ G TPLHIAA+ V +LLQ G S ++

Sbjct: 575 VAVHHNNLDIVKLLPRGSGSPHAWNGYTPLHIAAKQNQVEVARSLQYGGSAEVSQ 634

Query: 598 KETPLKCALNSKILSVMEAYHLSFERRQKSSEAPVQSPQRSVDSISOESSTSSFSM-SA 656  
TPL A M A LS +Q + +S + ++QE +

Sbjct: 635 GVTPLHLAAQEGHAE-MVALLS---KQANGNLGNKSGLTPLHLVAQEGHVPVADVLIK 690

Query: 657 GSRQEETKKDYREVEKLLRAVADGDLEMVRYLLEWTEEDLEDAEDTSAADPEFCHPLCQ 716  
G + T + L A G++++V++LL+ + D+ +A+ + + PL Q

Sbjct: 691 GVMVDATTR--MGYTPLHVASHYGNIKLVKFLQH-QADV-NAKTKLGS-----PLHQ 740

Query: 717 CPKCAPAQKRLAKVPASGLGVNVTSQDGSPLHVA 751  
+ + + +G N S DG++PL +A

Sbjct: 741 AAQQGHTDI-VTLLKNGASPNVSSDGTTPLAIA 774

Score = 162 (24.3 bits), Expect = 1.8e-07, Sum P(2) = 1.8e-07  
Identities = 48/149 (32%), Positives = 71/149 (47%)

Query: 737 VNVTSQDGSPLHVAALHGRADLIRLLKHNAGARNADQAVPLHLACQQGHFQVVKCL 796  
V D ++ AA G D L++G + N + LHLA ++GH ++V L

Sbjct: 5 VGFREDAATSFLRAARSGNLDKALDHLRNGVDINTCNQNGNLGLHLASKEGHVKMVVEL 64

Query: 797 LDSNAKPNKKDLSGNTPLIYACSGGHELVALLQHGASINASNKNTALHEAVIEKHV 856  
L GNT L A G E+V L+ +GA++NA + KG T L+ A E H+

Sbjct: 65 LHKEIILETTTCKGNTALHIAALAGQDEVVRELVNYGANVNAQSQKGFPLYMAAQENHL 124

Query: 857 FVVELLLLHGASVQVLNKRQRTAVDCAEQ 885  
VV+ LL +GA+ V + T + A Q

Sbjct: 125 EVVKFLENGANQNVATEDGFTPLAVALQ 153

Score = 158 (23.7 bits), Expect = 5.7e-26, Sum P(2) = 5.7e-26  
Identities = 38/135 (28%), Positives = 65/135 (48%)

Query: 460 DRRGHTPLHVAACGQASLIDLLVSKGAMVNATDYHGATPLHLACQKGYQSVTLLLLHYK 519  
+ G LH+A+ G ++ L+ K ++ T G T LH+A G V L++Y

Sbjct: 42 NQNGNLGLHLASKEGHVKMVVELLHKEIILETTTCKGNTALHIAALAGQDEVVRELVNYG 101

Query: 520 ASAEVQDNNGNTPLHLACTYGHEDCVKALVYDVESCRLDIGNEKGDTPHIAARWGYQ 579  
A+ Q G TPL++A H + VK L+ ++ E G TPL +A + G++

Sbjct: 102 ANVNAQSQKGFPLYMAAQENHLEVVKFLE---NGANQNVATEDGFTPLAVALQOQHEN 158

Query: 580 VIETLLQNGASTEIQ 594  
V+ L+ G ++

Sbjct: 159 VVAHLINYGTKGVR 173

Score = 115 (17.3 bits), Expect = 1.8e-21, Sum P(2) = 1.8e-21  
 Identities = 37/119 (31%), Positives = 58/119 (48%)

Query: 497 ATPLHLACQKGYQSVTL LLLHYKASAEVQ--DNNGNTPLHLACTYGHEDCVKALVYVDVE 554  
 AT A + G ++ L H + ++ + NG LHLA GH V L++ ++  
 Sbjct: 13 ATSFRLAARSG--NLDKALDHLRNGVDINTCNQNGNLGLHLASKEGHVVMVVELLHKEII 70

Query: 555 SCRLDIGNEKGDTPHLHIAARWGYQGVETLLQNGASTEIQNRLKETPLKCALNSKILSV 614  
 L+ +KG+T LHIAA G V+ L+ GA+ Q++ TPL A L V+  
 Sbjct: 71 ---LETTTKKGNTALHIAALAGQDEVVREL VNYGANVNAQSQKGF TPLYMAAQENHLEV 127

Query: 615 E 615  
 +  
 Sbjct: 128 K 128

Score = 106 (15.9 bits), Expect = 1.8e-01, Sum P(2) = 1.6e-01  
 Identities = 34/121 (28%), Positives = 54/121 (44%)

Query: 769 NAGARNADQAVPLHLACQGHFQVVKLLDSNAKPNKDLSGNTPLIYACSGGHHELVAL 828  
 + G R AD A A + G+ L + N + +G L A GH ++V  
 Sbjct: 4 SVGFREADAATSFLRAARSGNLDKALDHLRNGVDINTCNQNGNLGLHLASKEGHVVMVVE 63

Query: 829 LLQHGASINASNKNGNTALHEAVIEKHVVFVVELLLHGASVQVLNKRQRTAVDCAEQNSK 888  
 LL + + KGNTALH A + VV L+ +GA+V +++ T + A Q +  
 Sbjct: 64 LLHKEIILETTTKKGNTALHIAALAGQDEVVREL VNYGANVNAQSQKGF TPLYMAAQENH 123

Query: 889 I 889  
 +  
 Sbjct: 124 L 124

Score = 40 (6.0 bits), Expect = 1.6e-14, Sum P(2) = 1.6e-14  
 Identities = 11/56 (19%), Positives = 23/56 (41%)

Query: 622 ERRQKSSEAPVQSPQRSVDSISQESSTSSFSMSAGSROEETKKDYREVEKLLRAV 677  
 +RRQ+ E VQ + + + Q + + Q ++ +K++R V  
 Sbjct: 1614 DRRQQGQEEQVQEAKNFTTQVVGNEFQNIPEGEQVTEEQFTDEQGNIVTKKIIRKV 1669

Score = 38 (5.7 bits), Expect = 2.6e-14, Sum P(2) = 2.6e-14  
 Identities = 6/12 (50%), Positives = 10/12 (83%)

Query: 806 KDSGNTPLIYA 817  
 +D++G T L+YA  
 Sbjct: 1186 EDITGTTKL VYA 1197

Pedant information for DKFZphtes3\_1817, frame 2

#### Report for DKFZphtes3\_1817.2

[LENGTH] 1050  
 [MW] 117013.72  
 [pI] 6.47  
 [HOMOL] TREMBL:DMANKY\_1 product: "ankyrin"; Drosophila melanogaster ankyrin mRNA,  
 complete cds. 2e-45  
 [FUNCAT] 08.19 cellular import [S. cerevisiae, YOR034c] 5e-13  
 [FUNCAT] 10.05.99 other pheromone response activities [S. cerevisiae, YDR264c]  
 3e-12  
 [FUNCAT] 03.07 pheromone response, mating-type determination, sex-specific proteins  
 [S. cerevisiae, YDR264c] 3e-12  
 [FUNCAT] 99 unclassified proteins [S. cerevisiae, YIL112w] 2e-11  
 [FUNCAT] 06.13.01 cytoplasmic degradation [S. cerevisiae, YGR232w] 8e-10  
 [FUNCAT] 30.10 nuclear organization [S. cerevisiae, YIR033w] 2e-08  
 [FUNCAT] 04.05.01.07 chromatin modification [S. cerevisiae, YIR033w] 2e-08  
 [FUNCAT] 01.04.04 regulation of phosphate utilization [S. cerevisiae, YGR233c]  
 3e-08  
 [FUNCAT] 08.13 vacuolar transport [S. cerevisiae, YML097c] 5e-05  
 [FUNCAT] 06.04 protein targeting, sorting and translocation [S. cerevisiae, YML097c]  
 5e-05  
 [FUNCAT] 30.03 organization of cytoplasm [S. cerevisiae, YML097c] 5e-05  
 [FUNCAT] 08.07 vesicular transport (golgi network, etc.) [S. cerevisiae, YML097c]  
 5e-05  
 [FUNCAT] 03.22 cell cycle control and mitosis [S. cerevisiae, YER111c] 3e-04  
 [FUNCAT] 04.05.01.04 transcriptional control [S. cerevisiae, YER111c] 3e-04  
 [BLOCKS] BL00901A Cysteine synthase/cystathionine beta-synthase P-phosphate att  
 dlawcb 1.91.3.1.2 GA binding protein (GABP) alpha GA bindini 4e-12  
 [SCOP] 3.1.3.53 Myosin-light-chain-phosphatase 1e-12  
 [EC] phosphotransferase 1e-19  
 [PIRKW] nucleus 1e-13

[PIRKW] potassium channel 5e-15  
 [PIRKW] early protein 2e-13  
 [PIRKW] tumor suppressor 1e-09  
 [PIRKW] duplication 1e-14  
 [PIRKW] tandem repeat 1e-19  
 [PIRKW] heterodimer 1e-14  
 [PIRKW] potassium transport 5e-15  
 [PIRKW] cell cycle control 1e-10  
 [PIRKW] serine/threonine-specific protein kinase 1e-19  
 [PIRKW] transmembrane protein 5e-15  
 [PIRKW] transport protein 5e-15  
 [PIRKW] DNA binding 2e-11  
 [PIRKW] oncogene 1e-08  
 [PIRKW] ATP 1e-19  
 [PIRKW] protein kinase inhibitor 1e-09  
 [PIRKW] voltage-gated ion channel 5e-15  
 [PIRKW] phosphoprotein 4e-38  
 [PIRKW] apoptosis 1e-19  
 [PIRKW] liver 4e-09  
 [PIRKW] integrin binding 3e-16  
 [PIRKW] differentiation 2e-12  
 [PIRKW] transforming protein 1e-08  
 [PIRKW] alternative splicing 1e-40  
 [PIRKW] coiled coil 1e-14  
 [PIRKW] peripheral membrane protein 2e-38  
 [PIRKW] transcription factor 4e-16  
 [PIRKW] transcription regulation 2e-16  
 [PIRKW] nucleotide binding 5e-15  
 [PIRKW] phosphoric monoester hydrolase 1e-12  
 [PIRKW] cytoskeleton 8e-39  
 [PIRKW] calmodulin binding 1e-19  
 [PIRKW] smooth muscle 1e-12  
 [SUPFAM] ankyrin 1e-40  
 [SUPFAM] death-associated protein kinase 1e-19  
 [SUPFAM] ankyrin repeat homology 1e-40  
 [SUPFAM] protein kinase homology 1e-19  
 [SUPFAM] vaccinia virus 27.4K HindIII-C protein homology 3e-07  
 [SUPFAM] int-3 transforming protein 1e-08  
 [SUPFAM] unassigned ankyrin repeat proteins 2e-38  
 [SUPFAM] notch protein 2e-12  
 [SUPFAM] fowlpox virus BamHI-ORF7 protein 2e-13  
 [SUPFAM] rel homology 2e-11  
 [SUPFAM] EGF homology 2e-12  
 [PROSITE] ATP\_GTP\_A 1  
 [PFAM] Ank repeat  
 [KW] Irregular  
 [KW] 3D  
 [KW] LOW\_COMPLEXITY 3.05 %

SEQ MALYDELLKNPFYALQKCRPDLCCKVAQIHGIVLPCKGSLSSSIQSTCFESYILIP  
 SEG .....  
 lawcB .....  
 SEQ VEEHFQTLNGKDVFIQGNRIKLGAGFACLLSVPILFETTFYNEKEESFSLCIAHPLEKR  
 SEG .....  
 lawcB .....  
 SEQ ESSEELAPSDPFSLKTIEDVREFLGRHSEFDRNIASFHRTFRECERKSLRHHIDSANA  
 SEG .....  
 lawcB .....  
 SEQ LYTCKLQQLLRDHLKMLAKQEAQMNLKQAVEIYVHHEIYNLIFKYVGTMEASEDAAFN  
 SEG .....  
 lawcB .....  
 SEQ KITRSLQDLQKQDIGVKPEFSFNI PRAKRELAQLNKCTSPQQLVCLRKVVQLITQSPSQ  
 SEG .....  
 lawcB .....  
 SEQ RVNLETMCADDLLSVLLYLLVKTEIPNWMANLSYIKNFRFSSSLAKDELGYCLTSFEAAIE  
 SEG .....  
 lawcB .....  
 SEQ YIRQGSLSAKPPESEGFCDRLFLKQMSLLSMTSSPTDCLFKHIASGNQKEVERLLSQE  
 SEG .....  
 lawcB .....  
 SEQ DHDKDTVQKMCHPLCFDDCEKLVSGRLNDPSVVTFPSRDDRGHTPLHVAACGGQASLID  
 SEG .....  
 lawcB .....



```

SEQ      LLVSKGAMVNATDYHGATPLHLACQKGYQSVTL LLLHYKASAEVDNNGNTPLHLACTYG
SEG      .....
lawcB    .....

SEQ      HEDCVKALVYYDVESCRDLIGNEKGDTPHIAARWGYQGVETLLQNGASTEIQNRLKET
SEG      .....
lawcB    .....

SEQ      PLKCALNSKILSVMEAYHLSFERRQKSSEAPVQSPQRSVDSISQESSTSSFSSMSAGSRQ
SEG      .....XXXXXXXXXXXXXXXXXXXXX.
lawcB    .....

SEQ      EETKKDYREVEKLLRAVADGDLEMVRYLLEWTEEDLEDAEDTVSAADPEFCHPLCQCPKC
SEG      .....
lawcB    .....

SEQ      APAQKRLAKVPASGLGVNVTSDGSSPLHVAALHGRADLIRLLKHGANAGARNADQAVP
SEG      .....
lawcB    .....CHHHHHHHHHHCHHHHHHHHHHCCCC-CCTTTTCCCH

SEQ      LHLACQGHFQVVKCLLDSNAKPNKKDLSGNTPLIYACSGGHHELVALLQHGASINASN
SEG      .....
lawcB    HHHHHHHCHHHHHHHHHHCCCTTTTCTTTTCHHHHHHHHTHHHHHHHHHCCCTTTTEE

SEQ      NKGNTALHEAVIEKHVFVVELLLHGASVQVLNKRQRTAVDCAEQNSKIMELLQVVPSCV
SEG      .....
lawcB    TTTEHHHHHHHCHHHHHHHHHHCCCTTTTCBTTTBCHHHHHHHHCHHHHHC.....

SEQ      ASLDVAETDRKEYVTVKIRKWNKSLYDLPDEPFTQFYFVHSAGQFKGKTSREIMARD
SEG      .....
lawcB    .....

SEQ      RSVPNLTGSLHEPGRQSVTLRQNNLPAQSGSHAAEKGNSDWPERPLTQTGPGHRRMLR
SEG      .....
lawcB    .....

SEQ      RHTVEDAVVSQGPEAAGPLSTPQEVSAARS
SEG      .....
lawcB    .....

```

## Prosites for DKFZphtes3\_1817.2

PS00017    945->953    ATP\_GTP\_A    PDOC00017

## Pfam for DKFZphtes3\_1817.2

```

HMM_NAME      Ank repeat
HMM            *GyTPLHIAARYNNvEMVrILLQHGADIN*
               G+TPLH+AA ++ +++LL+++GA +N
Query          463  GHTPLHVAAVCGQASLIDLLVSKGAMVN      490

32.12 (bits) f: 496 t: 523 Target: dkfzphes3_1817.2 similarity to ankyrins
Alignment to HMM consensus:
Query          *GyTPLHIAARYNNvEMVrILLQHGADIN*
               G TPLH+A++ + ++ LLL + A+
dkfzphes3      496  GATPLHLACQKGYQSVTL LLLHYKASAE      523

Query          f: 529 t: 556 Target: dkfzphes3_1817.2 similarity to ankyrins
Alignment to HMM consensus:
HMM            *GyTPLHIAARYNNvEMVrILLQHGADIN*
               G+TPLH+A+ Y+++++V+ L+ +
Query          529  GNTPLHLACTYGHEDCVKALVYYDVESC      556

42.65 (bits) f: 565 t: 592 Target: dkfzphes3_1817.2 similarity to ankyrins
Alignment to HMM consensus:
Query          *GyTPLHIAARYNNvEMVrILLQHGADIN*
               G+TPLHIAAR + +++ LLQ+GA+
dkfzphes3      565  GDTPLHIAARWGYQGVETLLQNGASTE      592

Query          f: 744 t: 771 Target: dkfzphes3_1817.2 similarity to ankyrins
Alignment to HMM consensus:
HMM            *GyTPLHIAARYNNvEMVrILLQHGADIN*
               G +PLH+AA +++ +RRLL+HGA+
Query          744  GSSPLHVAALHGRADLIRLLKHGANAG      771

```

36.38 (bits) f: 777 t: 804 Target: dkfzphtes3\_1817.2 similarity to ankyrins  
Alignment to HMM consensus:  
Query \*GyTPLHIAARYNNvEMVrILLQHGADIN\*  
PLH+A++++ ++V+ LL+ +A +N  
dkfzphtes3 777 QAVPLHLACQGGHFQVVKCLLDSNAKPN 804

Query f: 810 t: 837 Target: dkfzphtes3\_1817.2 similarity to ankyrins  
Alignment to HMM consensus:  
HMM \*GyTPLHIAARYNNvEMVrILLQHGADIN\*  
G+TPL++A+ ++ E+V LLLQHG+IN  
Query 810 GNTPLIYACSGGHHELVALLQHGASIN 837

44.62 (bits) f: 843 t: 870 Target: dkfzphtes3\_1817.2 similarity to ankyrins  
Alignment to HMM consensus:  
Query \*GyTPLHIAARYNNvEMVrILLQHGADIN\*  
G+T+LH A+++ +V +V+LLL HGA++  
dkfzphtes3 843 GNTALHEAVIEKHVFVVELLLHGASVQ 870

DKFZphtes3\_19f19  
-----

group: testes derived

DKFZphtes3\_19f19 encodes a novel 254 amino acid protein with weak similarity to *S. cerevisiae* protein YFL046w.

The protein contains a RGD cell attachment site.  
No informative BLAST results; No predictive prosite, pfam or SCOP motif.

The new protein can find application in studying the expression profile of testis-specific genes.

similarity to YFL046w

localisation: 3 STS match perfect but HS1292427 matches to chromosome 4

Sequenced by MediGenomix

Locus: /map="405.0/.3 cR from top of Chr11 linkage group"

Insert length: 1395 bp

Poly A stretch at pos. 1367, no polyadenylation signal found

```
1 GGGACCACGG TGGCGCCTGC GCTGGGAGGT GAGCTTGTGA CAGAGCGAAA
51 ACTACAATTC CCAGCATTCC TGTGGTGCCA GAACCTACCTT GCCCGAAAAGC
101 CTGTGCGAGA TTTACCCCGT CTTCCGCTC CTCGCCACCG GAAACTCTGT
151 AGGACATGAA TAGTCGCCAG GCTTGGCGGC TCTTCTCTC CCAAGGCAGA
201 GGAGATCGTT GGGTTTCAAG GCCCGCGGG CATTCTCGC CGGCCCTGCG
251 GAGAGAGTTC TCACTACCA CAACCAAGGA GGGATATGAT AGGCGGCCAG
301 TGGATATAAC TCCTTTAGAA CAAGGAAAT TAACCTTTGA TACCATGCA
351 TTGGTTGAGG ACTTGGAAAC TCATGGATT GACAAACAC AAGCAGAAAC
401 AATTGTATCA GCGTTAATCG CTTTATCAAA TGTCAGCTG GATACTATCT
451 ATAAAGAGAT GGTCACTCAA GCTCAACAG AAATAACACT ACAACAGCTA
501 ATGGGCTCATT TGGATGCTAT CAGGAAAGC ATGGCTATCC TAGAGAAAAG
551 TGAATTTGCA AATCTGAGAG CAGAGAAAT GAAATGAAA ATTGAATTAG
601 ACCAAGTTAA GCATCAACTA ATGCATGAAA CCAGTCCAGT CAGAGCAGAT
651 AATAAAGTGG ATATCAACTT AGAAAGGAGC AGAGTAAACG ATATGTTTAC
701 AGATCAAGAA AAGCAACTTA TGGAAACAC TACAGATTT ACAAAAAGCG
751 ATACTCAAA CAAAAGTATT ATTTAGAGA CCAGTAATAA AATTGACGCT
801 GAAATTGCTT CCTTAAAAAC ACTGATGGAA TCTAACAAAC TTGAGACAAT
851 TCGTTATCTT GCAGCTTCGG TGTTTACTTG CTTGGCAATA GCATTGGGAT
901 TTTTAGATT CTGGAAGTAG TATTAAATGCT CATCCTGCTG TGGCTGTTGG
951 CTTCTTAGAA CACCAACCG GGAGAGATT ACTTGAACA TTGTCAGTTG
1001 CAGCAAAAAT TTACTACACA AGATTATTCC AAGTGTATAC GGACTAAAAG
1051 AGGAAGTGTT TTAGAATGAG AAGAGATACT GTGCTTTAT TGTGTGTGTG
1101 TGAGTGCAGG TGTGTGCTT TATTATATTG AAAAGCTGTC ACTCAGACCT
1151 GGTTTGAGAT AGAAGAGCAT TTTGTCCTTT TGATAGTTAA TAGAAATTGA
1201 ACCAGAGTTT TCTTATGTTT GCTTGAACAG TTGTGTAAT CATACAGGAT
1251 TTTGTGGGTA TTGGTTGAAT ATTTGTAAC CATTCCCTAG CCTACATATT
1301 TATTACTGAA TTAACCTTCC TGATAACCAT TGCATAATTA CATTCTTCTA
1351 TAAATGAAA GATTATTACA ACAAAAAAAA AAAAAAAA AAAAA
```

## BLAST Results

Entry HS419346 from database EMBL:  
human STS WI-13569.  
Score = 2154, P = 8.6e-91, identities = 446/459

Entry HS1292427 from database EMBL:  
human STS SHGC-50338.  
Score = 1737, P = 7.2e-72, identities = 359/369

Entry HS253344 from database EMBL:  
human STS WI-13893.  
Score = 1578, P = 1.0e-64, identities = 358/397

## Medline entries

No Medline entry

## Peptide information for frame 3

-----  
ORF from 156 bp to 917 bp; peptide length: 254  
Category: similarity to unknown protein  
Classification: no clue  
Prosite motifs: RGD (15-18)

1 MNSRQAWRLF LSQGRGDRWV SRPRGHFSPA LRREFFTTTT KEGYDRRPVD  
51 ITPLEQRKLT FDTHALVQDL ETHGFDKTA ETIVSALTAL SNVSLDTIYK  
101 EMVTQAQQEI TVQQLMAHLD AIRKDMVILE KSEFANLRAE NEKMKIELDQ  
151 VKQQLMHETS RIRADNKLDI NLEERSVTDI FTDQEQQLME TTTEFTKKDT  
201 QTKSISETS NKIDAEIASL KTLMESNKLE TIRYLAASVF TCLALALGFY  
251 RFWK

## BLASTP hits

No BLASTP hits available

## Alert BLASTP hits for DKFZphtes3\_19f19, frame 3

SWISSPROT:YAN8\_SCHPO HYPOTHETICAL 24.6 KD PROTEIN C3H1.08 IN CHROMOSOME  
I., N = 1, Score = 144, P = 8.4e-09

PIR:S56209 probable membrane protein YFL046w - yeast (Saccharomyces  
cerevisiae), N = 1, Score = 138, P = 5.4e-08

>SWISSPROT:YAN8\_SCHPO HYPOTHETICAL 24.6 KD PROTEIN C3H1.08 IN CHROMOSOME I.  
Length = 211

## HSPs:

Score = 144 (21.6 bits), Expect = 8.4e-09, P = 8.4e-09  
Identities = 34/121 (28%), Positives = 67/121 (55%)

Query: 70 LETHGFDKTAETIVSALTALSNVSLDTIYKEMVTQAQQE-ITVQQLMAHLDIAIRKDMVI 128  
LE G+ AETI + + ++ +L + K + +A+QE ++ QQ L IRK +  
Sbjct: 46 LEQAGYSVKNAETITNLMRTITGEALTELEKNIGFKAKQESVSFQQKRTFLQ-IRKYLET 104  
Query: 129 LEKSEFANLRAENEMKIELDQVKQQLMHETSRIADNKLDINLERSRVTDIEMTDQEQQL 188  
+E++EF +R ++K+ E+++ K L + ++ +L++NLE+ R+ D T + +  
Sbjct: 105 IEENEFDKVRKSSDKLINEIEKTKSSLRDVDKLTALSEVRLNLNLEKGRMKDAATSRNTNI 164  
Query: 189 ME 190  
E  
Sbjct: 165 HE 166

## Pedant information for DKFZphtes3\_19f19, frame 3

## Report for DKFZphtes3\_19f19.3

{LENGTH} 254  
{MW} 29505.73  
{pI} 6.99  
{HOMOL} PIR:S56209 probable membrane protein YFL046w - yeast (Saccharomyces cerevisiae)  
2e-10  
{FUNCAT} 99 unclassified proteins [S. cerevisiae, YFL046w] 8e-12  
{PROSITE} RGD 1  
{KW} TRANSMEMBRANE 1  
{KW} LOW\_COMPLEXITY 5.12 %  
{KW} COILED\_COIL 11.02 %

SEQ MNSRQAWRLF LSQGRGDRWVSRPRGHFSPA LRREFFTTTT KEGYDRRPVDITPLEQRKLT  
SEG .....  
PRD ccchhhhhhhhhccccccccccccchhhhhhhheeeccccccccccchhhhhhhcc  
COILS .....  
MEM .....  
SEQ FDTHALVQDLETHGFDKTAETIVSALTALSNVSLDTIYKEMVTQAQQEITVQQLMAHLD  
SEG .....  
PRD chhhhhhhhhhhccccchhh  
COILS .....

**PCT/IB00/01496**

Prosites for DKFZphtes3\_19f19.3

(No Pfam data available for DKFZphtes3\_19f19.3)

DKFZphtes3\_19j17  
-----

group: testes derived

DKFZphtes3\_19j17 encodes a novel 436 amino acid protein with partial similarity to C.elegans Y40B1A.2 protein.

The novel protein contains two Prosite WW/rsp5/WWP domain signatures. The WW domain (or rsp5 or WWP domain) has been originally discovered as a short conserved region in a number of unrelated proteins, such as dystrophin, utrophin, vertebrate YAP protein, mouse NEDD-4 and yeast RSP5. The domain is repeated up to 4 times in some proteins. It has been shown to bind proteins with particular proline-motifs, [AP]-P-P-[AP]-Y, and thus resembles somewhat SH3 domains. It appears to contain beta-strands grouped around four conserved aromatic positions; generally Trp. The name WW or WWP derives from the presence of these Trp as well as that of a conserved Pro. It is frequently associated with other domains typical for proteins in signal transduction processes.

No informative BLAST results; No predictive prosite, pfam or SCOP motive.

The new protein can find application in studying the expression profile of testis-specific genes.

similarity to C.elegans Y40B1A.2

there are two long ORFs in this cDNA according to EST:  
HS12146/HS75086/AA923755/MMAA17335 remaining intron at Bp 1506-1733

Sequenced by MediGenomix

Locus: unknown

Insert length: 2762 bp

Poly A stretch at pos. 2740, no polyadenylation signal found

```
1 ATTCTCAGCC AAATTTTTTT ATTTTGTGCA GAATCAGTGT GCAAGGTGGT
51 TTATAAGATA ATGGAGTGGT TTTTTTTTGT GTTTAGTGTG ATTTGTTATC
101 AGGAGTCTTA TTGTAACGCT TAAGCATTAG GTTTTTTGTG TGAGAAACCT
151 TAAAGAGTAA AGCAGAATTG AAAGTGGAAA TTTTAATTTT GTAAGTTCAT
201 AAAAATTTAA GATAATACAC CAAAGTTTAT GTTTAAATTA GGGAGTTTAA
251 GGTTCATATT CTTTCTCTTT TTTTTGGGG GGGTGATGTT TTACAGGAC
301 TTAAGTATTC ATCGAAGAGT CACCCACGTA GCGGTGATCA CAGACATGAA
351 AAGATGCGAG ACGCCGGAGA TCCTTCACCA CCAAAATAAA TGTTCGGAG
401 ATCTGATAGT CCTGAAAACA AATACAGTGA CAGCACATGT CACAGTAAAG
451 CCAAAAATGT GCATACTCAC AGAGTTAGAG AGAGCGATGG TGGGACCACT
501 TACTCTCCAC AAGAAAATTC ACACACCCAC AGTCTCTTTC ATAGTTCAAA
551 TTCACATTCT TCTAATCCAA GCATAACCC AAGCAAAACT TCAGATCCAC
601 CTTATGATTC TCGAGATGAC TCGTCTGAGC ATATTAGTTC TTCTGGGAAA
651 AAGTACTACT ACAATTGTCG ACAGAGAGTT TCACATGGG AAAAACCAAA
701 AGACTGGCTT GAAGAGAAC AGAGACAAA AGAGCAAAAC AAGATGGCAG
751 TCACACAGCT CCCAAAGAT AGGGATTACA GAAGAGAGGT GATGCAAGCA
801 ACAGCCACTA GTGGGTTTGC CAGTGGAAAT GAAGACAAGC ATTCCAGTGA
851 TGCCAGTAGT TTGCTCCAC AGAATATTTT GTCTCAAAAC AGCAGACACA
901 ATGACAGAGA CTACAGACTG CCAAGAGCAG AGACTCACAG TAGTTCTACG
951 CAGTACAGC ACCCCATCAA ACCAGTGGTT CATCCAACTG CTACCCCAAG
1001 CACTGTTCTT TCTAGTCCAT TTACGCTACA GTCTGATCAC CAGCCAAAGA
1051 AATCATTGTA TGCTAATGGA GCATCTACTT TATCAAACT GCCTACACCC
1101 ACATCTTCTG TCCCTGCACA GAAACAGAAA AGAAAAGAAT CTACATCAGG
1151 AGACAAACCC GTATCACATT CTTGCACAC TCCTTCCAGC TCTTCTGCCCT
1201 CTGACTGAA CCCACATCT GCACCTCCAA CATCTGCTTC AGCGGTCCCT
1251 GTTCTCCTG TTCCACAGTC GCCAATACCT CCCTTACTTC AGGACCCAAA
1301 TCTTCTTAGA CAATTGCTTC CTGCTTTGCA AGCCACGCTG CAGCTTAATA
1351 ATTCTAATGT GGACATATCT AAAATAAATG AAGTTCTTAC AGCAGCTGTG
1401 ACACAAGCCT CACTGCAGTC TATAATTCAT AAGTTCTTTC CTGCTGGACC
1451 ATCTGCTTTC AACATAACGT CTCTGATTTC TCAAGTGTCT CAGCTCTCTA
1501 CACAAGATAT CCCTCTTCAT GAAGGTATCC AAATGGAGAG AGATACACAT
1551 AGGAGCAAAAT GGAAGTGAA AGGGTCACCT TGTGAGAAAG CTGATAACAA
1601 GCAGGAATGC CTTGCTGGA ATGGAAGTAT AATGGTGCAA AGACTCTTGC
1651 AACCCTCTGG CTAGCTCAT GAGCAGGAGA CTGCGTGGGA TACCTGGGCC
1701 TAAATGTAGA ATAAGAAAGA AGAAATAAGG ATGCCAGGCC ATCTAATCAG
1751 TCTCCGATGT CTTTAACATC TGATGCGTCA TCCCAAGATC CATATGTTTC
1801 TCCAAGAATA AGCACACCTC AAACCTAACAC AGTCCCTATC AAACCTTTGA
1851 TCAGTACTCC TCCTGTTTCA TCACAGCCAA AGGTTAGTAC TCCAGTAGTT
1901 AAGCAAGGAC CAGTGTACCA GTCAGCCACA CAGCAGCCTG TAACTGCTGA
1951 CAAGCAGCAA GGTATGAAC CTGTCTCTCC TCGAAGCTTT CAGCGCTCAA
2001 GCCAGAGAAG TCCATCACCT GGTCCCAATC ATACTTCTAA TAGTAGTAAT
2051 GCATCAAAATG CAACAGTTGT ACCACAGAAAT TCTTCTGCCG GATCCACGTG
```

```
2101 TTCATTAACG CCTGCACTAG CAGCACACTT CAGTGAAAAT CTCATAAAAC
2151 ACGTTCAAGG ATGGCCTGCA GATCATGCAG AGAAGCAGGC ATCAAGATTA
2201 CGCGAAGAAG CGCATAACAT GGGAACTATT CACATGTCCG AAATTGTGAC
2251 TGAATTAAAA AATTTAAGAT CTTTAGTCCG AGTATGTGAA ATCAAGCAA
2301 CTTTGCGAGA GCAAAGGATA CTATTTTGA GACAACAAAT TAAGGAACCT
2351 GAAAAGCTAA AAAATCAGAA TTCCTTCATG GTGTGAAGAT GTGAATAATT
2401 GCACATGGTT TTGAGAACAG GAACTGTAAA TCTGTTGCC AATCTTAACA
2451 TTTTGTAGCT GCATTTAAGT AGACTTTGGA CCGTTAAGCT GGGCAAAGGA
2501 AATGACAAGG GGACGGGGTC TGTGAGATC AATTCAGGGG AAAGATACAA
2551 GATTGATTG TAAAACCCCTT GAAATGTAGA TTTCTGTAG ATGTATCCTT
2601 CACGTTGTAA ATATGTTTTG TAGAGTGAAG CCATGGGAAG CCATGTGTAA
2651 CAGAGCTTAG ACATCCAAAA CTAATCAATG CTGAGGTGGC TAAATACCTA
2701 GCCTTTTACA GTTAAACCTG TCTGCAAAAT TAGCTTTTTT AAAAAAAAAA
2751 AAAAAAAAAA AA
```

## BLAST Results

Entry AC005876 from database EMBLNEW:  
Homo sapiens chromosome 10 clone CIT987SK-118815 map 10p11.2-10p12.1,  
complete sequence.  
Score = 2130, P = 0.0e+00, identities = 426/426  
12 exons matching Bp 492-2740

## Medline entries

No Medline entry

## Peptide information for frame 2

ORF from 1757 bp to 2383 bp; peptide length: 209  
Category: questionable ORF  
Classification: no clue

```
1 MSLTSDASSP RSYVSPRIST PQNTVPKPK LISTPPVSSQ PKVSTPVVKQ
51 GPVSQATQQ PVTADKQGGH EPVSPRLQR SSQSPSPGP NHTSNSSNAS
101 NATVVPQNSS ARSTCSLTPA LAAHFSENLI KHVQGWFPADH AEKQASRLRE
151 EAHNMGTIHM SEICTELKNL RSLRVCEIQ ATLREQRILE LRQIQELEK
201 LKNQNSFMV
```

## BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphtes3\_19j17, frame 2

No Alert BLASTP hits found

## Peptide information for frame 3

ORF from 354 bp to 1661 bp; peptide length: 436  
Category: similarity to unknown protein  
Classification: unclassified  
Prosites motifs: WW DOMAIN\_1 (90-116)  
WW\_DOMAIN\_1 (90-116)

```
1 MRDAGDPSPP NKMLRRSDSP ENKYSDSTGH SKAKNVHTR VRERDGGTSY
51 SPOENSHNHS ALHSSNSHSS NPSNNPSKTS DAPYDSADDW SEHISSSGKK
101 YYNCRTEVS QWEKPKEWLE REQROKEANK MAVNSFPKDR DYRREVMQAT
151 ATSGFASGME DKHSSDASSL LPQNILSOTS RHNDRDYRLP RAETHSSSTP
201 VQHPIKPVVH PTATPSTVPS SPFTLQSDHQ PKKSFANGA STLKSLPTPT
251 SSVPAQKTER KESTSGDKPV SHSCTTPSTS SASGLNPTSA PPTSASAVPV
301 SPVQSPPIPP LLQDPNLLRQ LLPALQATLQ LNNSNVDISK INEVLTAATV
351 QASLQSIHK FLTAGPSAFN ITSLSISQAAQ LSTQDIPLHE GIQMERDTHR
401 SKWEVKGSLC QKADKQCECL VWNGSIMVQR LLQPSG
```

## BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphtes3\_19j17, frame 3

TREMBL:CEY40B1A\_2 gene: "Y40B1A.2"; Caenorhabditis elegans cosmid  
Y40B1A, N = 1, Score = 144, P = 1.8e-09

>TREMBL:CEY40B1A\_2 gene: "Y40B1A.2"; Caenorhabditis elegans cosmid Y40B1A  
Length = 120

HSPs:

Score = 144 (21.6 bits), Expect = 1.8e-09, P = 1.8e-09  
Identities = 30/67 (44%), Positives = 43/67 (64%)

Query: 90 WSEHISSSGKKYYNCRTEVSQWEKPKEW-LEREQRQKEANKMAVNSFPK---DRDYRRE 145  
W+E +SSSGK YYN +TE+SQW+KP EW E +++ K VN P+ DR Y  
Sbjct: 11 WTEQMSSSGKMYYNKKTEISQWDKPAEWPAEGGSAERDKPKGGVNEKPRFAEDR-YNEY 69

Query: 146 VMQATATS 153  
+ Q +++S  
Sbjct: 70 IGQLSSSS 77

Pedant information for DKFZphtes3\_19j17, frame 2

Report for DKFZphtes3\_19j17.2

[LENGTH]	209
[MW]	22873.85
[pI]	9.95
[KW]	All_Alpha
[RW]	LOW_COMPLEXITY 13.40 %

SEQ	MSLTSDASSPRSVSPRISTPQNTVPIKPLISTPPVSSQPKVSTPVVKQGPVVSQSATQQ
SEG	.....
PRD	cc
SEQ	PVTADKQGGHEPVSPRSLQRSSQSPSPGPNHTSNSSNASHATVVPQNSSARSTCSLTPA
SEG	.....XX
PRD	cc
SEQ	LAHFSENLIKHHVQGWADHAQASRLREEAHNMGTIMHSEICTELKNLRLVVRVCEIQ
SEG	.....
PRD	hhhhhhcc
SEQ	ATLREQRILFLRQQIKELEKLNQNSFMV
SEG	.....
PRD	hh

(No Prosite data available for DKFZphtes3\_19j17.2)

(No Pfam data available for DKFZphtes3\_19j17.2)

Pedant information for DKFZphtes3\_19j17, frame 3

Report for DKFZphtes3\_19j17.3

[LENGTH]	436
[MW]	47716.62
[pI]	8.71
[HOMOL]	TREMBL:CEY40B1A_2 gene: "Y40B1A.2"; Caenorhabditis elegans cosmid Y40B1A 6e-08
[FUNCAT]	04.05.03 mrna processing (splicing) [S. cerevisiae, YKL012w] 2e-04
[FUNCAT]	30.10 nuclear organization [S. cerevisiae, YKL012w] 2e-04
[FUNCAT]	99 unclassified proteins [S. cerevisiae, YPR152c] 6e-04
[BLOCKS]	BL01159 WW/rsp5/WWP domain proteins
[PROSITE]	WW DOMAIN 1 2
[PFAM]	WW/rsp5/WWP domain containing proteins
[KW]	All_Alpha
[RW]	LOW_COMPLEXITY 22.48 %



```

SEQ      MRDAGDPSPPNKMLRRSDSPENKYS DSTGHSKAKNVHTRVREDDGTSYSPQENSHNHS
SEG      .....
PRD      ccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc

```

```

SEQ      ALHSSNSHSSNPNPSKTS DAPYDSADDWSEHISSSGKKYYNCRTEVSQWEKPKEWLE
SEG      xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx
PRD      cccccccccccccccccccccccccccccccccccccccccccccccccccccccccchhh

```

```

SEQ      REQRQKEANKMAVNSFPKORDYRREVMQATATSGFASGMEDKHSSDASSLLPQNILSQT
SEG      .....
PRD      hhhhhhhhhhhhhccccccccchhhhhhhhhhhcccccccccccccccccccccccccccccc

```

```

SEQ      RHNDRDYRLPRAETHSSSTPVQHPKPVVHPTATPSTVPSSPFTLQSDHQP KKSFDANGA
SEG      .....
PRD      ccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc

```

```

SEQ      STL SKLPTPTSSVPAQKTERKESTSGDKPVSHSCTTPSTSSASGLNPTSAPPTSASAVPV
SEG      xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx
PRD      ccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc

```

```

SEQ      SPVPQSPIPFLQDPNLLRQLPALQATLQLNNSNVDISKINEVLTAAVTQASLQSI IHK
SEG      xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx
PRD      cccccccccccccccccchhhhhhhhhhhhhccccccccchhhhhhhhhhhhhhhhhhhhh

```

```

SEQ      FLTAGPSAFNITSLISQAAQLSTQDIP LHEGIQMERDTHRSKWEVKGSLCQKADKQEQECL
SEG      .....
PRD      hhccccccceehhhhhhhhhhhccccccccccccccccccccccccccccccccchhhhhhhccce

```

```

SEQ      VMNGSIMVQRLQPSG
SEG      .....
PRD      eecchhhhhhhcccc

```

## Prosite for DKFZphtes3\_19j17.3

PS01159	90->116	WW_DOMAIN_1	PDOC50020
PS01159	90->116	WW_DOMAIN_1	PDOC50020

## Pfam for DKFZphtes3\_19j17.3

```

HMM_NAME      WW/rsp5/WWP domain containing proteins
HMM            *LPaGWEeHWOpSGRpWYYWNHETkTQWEpP*
               + **W EH+* SG+ YY+N T+ *QWE+P
Query         86 SADDWSEHISSSGKK-YYNCRTEVSQWEKP 115

```

DKFZphtes3\_lcl

group: signal transduction

DKFZphtes3\_lcl encodes a novel 632 amino acid putative GTPase-activating protein, related to drosophila rotund transcript and human n-chimaerin.

rac small GTPase is associated with type-I phosphatidylinositol 4-phosphate 5-kinase and regulating the production of phosphatidylinositol 4,5-bisphosphate. The new protein is expected to activate p21rac-related small GTPases.

The new protein can find application in modulating/blocking the response to a cellular receptor.

similarity to GTPase-activating proteins

complete cDNA, complete cds, EST hits

Sequenced by DKF2

Locus: unknown

Insert length: 3237 bp

Poly A stretch at pos. 3227, no polyadenylation signal found

```
1 GCGAAGTGAA GGGTGGCCCA GGTGGGGCCA GGCTGACTGA ATGTATCTCC
51 TAGCTATGGA CTAATAATA CATGGGGGGA AATAACAAG TATTCATGAG
101 GGTGAAAATG TGACCCAGCA GGAAAATTAC AACTATTTTC AATTGACGTT
151 GAATAGGATG AGTCATGGAA TTTAAGTGAT TTAAGTGAAG TTATACTACT
201 GGTAGATAGA AGAGCTAAAG AAAGATGGAT ACTATGATCG TGAATGTGCG
251 GAATCTGTTT GAGCAGCTTG TGCGCCGGGT GGAGATTCTC AGTGAAGGAA
301 ATGAAGTCCA ATTTATCCAG TTGGCGAAGG ACTTTGAGGA TTTCGGTAAA
351 AAGTGGCAGA GGAAGTACCA TGAGCTGGGG AAATACAAGG ATCTTTTGAT
401 GAAAGCAGAG ACTGAGCGAA GTGCTCTGGA TGTTAAGCTG AAGCATGCAC
451 GTAATCAGGT GGATGTAGAG ATCAAACGGA GACAGAGAGC TGAGGCTGAC
501 TGCGAAAAGC TGAACGACA GATTCAGCTG ATTCGAGAGA TGCTCATGTG
551 TGACACATCT GGCAGCATTG AACTAAGCGA GGAGCAAAAA TCAGCTCTGG
601 CTTTCTCAA CAGAGGCCAA CCATCCAGCA GCAATGCTGG GAACAAAAAG
651 CTATCAACCA TTGATGAATC TGGTTCCATT TTATCAGATA TCAGCTTTGA
701 CAAGACTGAT GAATCACTGG ATTGGGACTC TTCTTTGGTG AAGACTTTGA
751 AACTGAAAGAA GAGAGAAAAG AGGGGCTCTA CTAGCCGACA GTTTGTTGAT
801 GGTCCCGCTG GACCTGTAAA GAAAACCTCT TCCATGGGCT GTTGGTGTAG
851 CCAGGGGAAAT GAATCCATAG TTGCAAAAC TACAGTGACT GTTCCCAATG
901 ATGGCGGGCC CATCGAAGCT GTCTCACTA TTGAGACTGT GCCATATTGG
951 ACCAGAGCC GAAAGAAAC AGGTACTTTA CAACCTTGGG ACAGTGACTC
1001 CACCCTGAAC AGCAGCGACC TGGAGCCAAG AACTGAGACA GACAGTGTGG
1051 GCAGCGCACA GAGTAATGGA GGGATGCGCC TCGATGACTT TGTTCCTAAG
1101 ACGGTTATTA AACCTGAATC CTGTGTTCCA TGTGGAAAGC GGATAAAATT
1151 TGGCAAAATTA TCTCTGAATC GTCGAGACTG TCGTGTGGTC TCTCATCCAG
1201 AATGTCGGGA CCGCTGTCCC CTTCCCTGCA TTCTACCCCT GATAGGAACA
1251 CCTGTCAAGA TTGGAGAGGG AATGCTGGCA GACTTTGTGT CCCAGACTTC
1301 TCCAATGATC CCCTCCATTG TTGTGCATTG TGTAAATGAG ATTGAGCAAA
1351 GAGGTCTGAC TGAGACAGGC CTGTATAGGA TCTCTGGCTG TGACCGCACA
1401 GTAAAAGAGC TGAAAGAGAA ATTCCTCAGA GTGAAAACCTG TACCCTCTCT
1451 CAGCAAGATG GATGATATCC ATGCTATCTG TAGCCCTCTA AAAGACTTTT
1501 TTCGAAACCT CAAAGAACCT CTTCTGACCT TTGCGCTTAA CAGAGCCTTT
1551 ATGGAAGCAG CAGAAATCAC AGATGAAGAC AACAGCATAG CTGCCATGTA
1601 CCAAGCTGTT GGTGAACCTG CCCAGGCCAA CAGGGACACA TTAGCTTTCC
1651 TCATGATTCA CTGACAGAGA GTGGCTCAGA GTCCACATAC TAAATGGAT
1701 GTTGCCAATC TGGCTAAAGT CTTTGGCCCT ACAATAGTGG CCCATGCTGT
1751 GCCCAATCCA GACCCAGTGA CAATGTTACA GGACATCAAG CGTCAACCCA
1801 AGGTGGTTGA GCGCCTGCTT TCCTTGCCCT TGGAGTATTG GAGTCAGTTT
1851 ATGATGGTGG AGCAAGAGAA CATTGACCCC TCATCATGTA TTGAAAACCT
1901 AAATGCCTTT TCAACACCAC AGACACCAGA TATTAAAGTG AGTTTACTGG
1951 GACCTGTGAC CACTCCTGAA CATCAGCTTC TCAAGACTCC TTCTCTAGT
2001 TCCCTGTGAC AGAGAGTCCG TTCCACCCTC ACCAAGAAAC CTCCTAGATT
2051 TGGGAGCAAA AGCAAGCTCT CCACTAACCT AGGACGACAA GGCAACTTTT
2101 TTGCTTCTCC AATGCTCAAG TGAAGTCACA TCTGCTGTT ACTTCCAGC
2151 ATTGACTGAC TATAAGAAAG GACACATCTG TACTCTGCTC TGCAGCCTCC
2201 TGTAATCATT ACTACTTTTA GCATTCTCCA GGCTTTTACT CAAGTTTAAT
2251 TGTCATGAG GGTTTTATTA AAATATATA TATCTCCCTC TCCTTCTCCT
2301 CAAGTCACAT AATATCAGCA CTTTGTGCTG GTCATTGTTG GGAGCTTTTA
2351 GATGAGACAT CTTTCCAGGG GTAGAAGGGT TAGTATGGAA TTGGTTGTGA
2401 TTCTTTTGG GGAAGGGGGT TATTGTTCTT TTGGCTTAAA GCCAAATGCT
2451 GCTCATAGAA TGATCTTTCT CTAGTTTCTT TTAGAAGCTA TTTCGGTGAG
2501 ACAATGACAG AAACCTTACC TATCTGATAA GATTAGCTTG TCTCAGGGTG
2551 GGAAGTGGGA GGGCAGGGCA AAGAAAGGAT TAGACCAGAG GATTAGGAT
```

```

2601 GCCTCCTTCT AAGAACCAAG AGTTCTCATT CCCCATTATG AACTGAGCTA
2651 TAATATGGAG CTTTCATAAA AATGGGATGC ATTGAGGACA GAACTAGTGA
2701 TGGGAGTATG CGTAGCTTTG ATTTGGATGA TTAGGTCTTT AATAGTGTGG
2751 AGTGGCACAA CTTGTAAAT GTGAAAGTAC AACTCGTATT TATCTCTGAT
2801 GTGCCGCTGG CTGAACCTTG GGTTCATTTG GGGTCAAGCC CAGTTTCTCT
2851 TTTAAATTTG AATTCACTCT GATGCTTGGC CCCCATACCC CCAACCTTGT
2901 CCACTGGAGC CCAACTTCTA AAGGTCAATA TATCATCCTT TGGCATCCCA
2951 ACTAACAATA AAGAGTAGGC TATAAGGGAA GATTGTCAAT ATTTTGTGGT
3001 AAGAAAAGCT ACAGTCATTT TTTCTTTGCA CTTTGGATGC TGAATTTTGT
3051 CCCATGGAAC ATAGCCACAT CTAGATAGAT GTGAGCTTTT TCTTCTGTGA
3101 AATTTATCTT TAATGCTCTG TAAACCGATT TTCTTCTGTA GAATGTTTGA
3151 CTTCTGTTTG ACCCTTATCT GTAAACACCC TATTTGGGAT AATATTGGA
3201 AAAAAAGTAA ATAGCTTTTT CAAAATGAAA AAAAAA

```

## BLAST Results

Entry U82984 from database EMBLEST:  
Homo sapiens DRES 56 mRNA sequence.  
Score = 8775, P = 0.0e+00, identities = 1757/1758  
matches 3' end

## Medline entries

93074974:  
Developmental regulation and neuronal expression of the mRNA of rat  
n-chimaerin, a  
p21rac GAP:cDNA sequence.

93024458:  
A Drosophila rotund transcript expressed during spermatogenesis and  
imaginal disc  
morphogenesis encodes a protein which is similar to human Rac  
GTPase-activating  
(racGAP) proteins.

## Peptide information for frame 3

ORF from 225 bp to 2120 bp; peptide length: 632  
Category: similarity to known protein

```

1 MDTHMLNVRN LFEQLVRRVE ILSEGNEVQF IOLAKDFEDF RKKWQRTDHE
51 LGKYKDLLMK AETERSALDV KLKHARNOVD VEIKRRQRAE ADCEKLERQI
101 QLIREMLMCD TSGSIOLSEE QKSALAFLENR GOPSSSNAGN KRLSTIDESG
151 SILSDISFDK TDESLWDSS LVKTFKLKKR EKRRSTSRQF VDGPPGPVKK
201 TRSIGSAVDQ GNEISIVAKTT VTVFNDGGPI EAVSTIETVP YWTRSRRTKG
251 TLOPWNDSST LNSRQLEPRT ETDSVGTQPS NGMRLHDFV SKTVIKPESC
301 VPCGKRIKFG KLSLKCRCRDR VVSHPECRDR CPLPCIPTLI GTPVKIGEGM
351 LADFSQTSPT MIPSIVVHCV NEIEQRGLTE TGLYRISGCD RTVKELKEKF
401 LRVKTVPLLS KVDDIHAICS LLKDFLRNLK EPLLTFRLENR AFMEAAEITD
451 EDNSIAAMYQ AVGELPQANR DTLAFLMIHL QRVAQSPHTK MDVANLAKVF
501 GPTIVAHAVP NPDPTMLQD IKRQPKVVER LLSLPLEYWS QFMVVEQENI
551 DPLHVIENSN AFSTPQTPDI KVSLLGPVTT PEHQLLKTFS SSSLSQVRVS
601 TLTKNTPRFG SKSKSATNLG RQGNFFASPM LK

```

## BLASTP hits

Entry CEK08E3\_4 from database TREMBLNEW:  
gene: "K08E3.6"; Caenorhabditis elegans cosmid K08E3  
Score = 452, P = 2.6e-48, identities = 126/377, positives = 189/377

Entry A48122 from database PIR:  
GTPase-activating protein Rac homolog, splice form clone pcl.7 - fruit  
fly (Drosophila melanogaster) (fragment)  
Score = 480, P = 9.2e-46, identities = 111/270, positives = 155/270

Entry B48122 from database PIR:  
GTPase-activating protein Rac homolog, splice form clone pcl.7d - fruit  
fly (Drosophila melanogaster)  
Score = 480, P = 9.2e-46, identities = 111/270, positives = 155/270

Entry DM22539.1 from database TREMBL:  
gene: "rotund"; product: "InracGAP"; *Drosophila melanogaster* rnracGAP  
(rotund) gene, complete cds.  
Score = 480, P = 9.2e-46, identities = 111/270, positives = 155/270

Entry S29128 from database PIR:  
N-chimerin - rat  
Score = 336, P = 8.8e-30, identities = 86/253, positives = 128/253

Alert BLASTP hits for DKFZphtes3\_1c1, frame 3

No Alert BLASTP hits found

Pedant information for DKFZphtes3\_1c1, frame 3

Report for DKFZphtes3\_1c1.3

[LENGTH] 632  
[MW] 71026.84  
[pI] 9.08  
[HOMOL] PIR:B48122 GTPase-activating protein Rac homolog, splice form clone pcl.7d -  
fruit fly (*Drosophila melanogaster*) 2e-46  
[FUNCAT] 10.99 other signal-transduction activities [S. cerevisiae, YBR260c] 3e-12  
[FUNCAT] 03.22 cell cycle control and mitosis [S. cerevisiae, YER155c] 2e-11  
[FUNCAT] 30.03 organization of cytoplasm [S. cerevisiae, YER155c] 2e-11  
[FUNCAT] 03.04 budding, cell polarity and filament formation [S. cerevisiae, YER155c]  
2e-11  
[FUNCAT] 03.10 sporulation and germination [S. cerevisiae, YDL240w] 3e-09  
[FUNCAT] 30.04 organization of cytoskeleton [S. cerevisiae, YOR134w] 4e-09  
[FUNCAT] 06.10 assembly of protein complexes [S. cerevisiae, YOR134w] 4e-09  
[FUNCAT] 03.07 pheromone response, mating-type determination, sex-specific proteins  
[S. cerevisiae, YOR127w] 5e-09  
[FUNCAT] 09.04 biogenesis of cytoskeleton [S. cerevisiae, YPL115c] 3e-08  
[FUNCAT] 10.02.09 regulation of g-protein activity [S. cerevisiae, YPL115c] 3e-08  
[BLOCKS] BL00479B Phorbol esters / diacylglycerol binding domain proteins  
[BLOCKS] BL00479A Phorbol esters / diacylglycerol binding domain proteins  
[SCOP] dlphwa\_1.83.1.1.2 p85 alpha subunit RhoGAP domain [human (Homo le-55  
[SCOP] dlrgp\_1.83.1.1.1 p50 RhoGAP domain [human (Homo sapiens) 1e-49  
[PIRKW] breakpoint cluster region 1e-19  
[PIRKW] transmembrane protein 7e-08  
[PIRKW] brain 3e-22  
[PIRKW] alternative splicing 1e-19  
[PIRKW] P-loop 2e-25  
[SUPFAM] CDC24 homology 3e-22  
[SUPFAM] bcr protein 3e-22  
[SUPFAM] myosin motor domain homology 2e-25  
[SUPFAM] pleckstrin repeat homology 4e-10  
[SUPFAM] LIM metal-binding repeat homology 2e-09  
[SUPFAM] protein kinase C zinc-binding repeat homology 5e-29  
[PROSITE] MYRISTYL 6  
[PROSITE] AMIDATION 1  
[PROSITE] CAMP\_PHOSPHO\_SITE 3  
[PROSITE] CK2\_PHOSPHO\_SITE 13  
[PROSITE] TYR\_PHOSPHO\_SITE 2  
[PROSITE] PKC\_PHOSPHO\_SITE 9  
[PROSITE] ASN\_GLYCOSYLATION 1  
[PROSITE] DAG\_PE\_BINDING\_DOMAIN 1  
[PFAM] Phorbol esters / diacylglycerol binding domain  
[KW] Irregular  
[KW] 3D  
[KW] LOW\_COMPLEXITY 2.22 %  
[KW] COILED\_COIL 8.54 %

SEQ MDTMMLNVRNLFELQVRRVEILSEGNEVQFIQLAKDFEDFRKKWQRTDHELKGYKDLLMK  
SEG .....  
COILS .....CCCCCCCCCCCC  
1rgp- .....  
SEQ AETERSALDVKLKHARNQVDVEIKRRQRAEADCEKLERQIQLIREMLMCDTSGSIQLSEE  
SEG .....  
COILS CC  
1rgp- .....  
SEQ QKSALAFNLNRGQPSSSNAGNKRLSTIDESGSILSDISFDKTDSESLDWDSSLVKTFKLKKR  
SEG .....  
COILS .....

```

1rgp- .....
SEQ      EKRRSTSRQFVDGPPGPVKKTRSIGSAVDQGNESIVAKTTVTVPNDGGPIEAVSTIETVP
SEG      .....
COILS    .....
1rgp- .....

SEQ      YWTRSRRTGTGLQPWNSDSTLNSRQLEPRTETDSVGT PQSNGGMRLHDFVSKTVIKPESC
SEG      .....
COILS    .....
1rgp- .....

SEQ      VPCGKRIRKFGKLSLKCRCRVVSHPECRDRCLPCIPTLIGTPVKIGEGMLADFSQTSP
SEG      .....
COILS    .....
1rgp- .....

SEQ      MIPSIVVHCVNEIEQRLTETGLYRISGCDRTVKELKEKFLRVKTVPLLSKVDDIHAICS
SEG      .....
COILS    .....
1rgp-    .CCHHHHHHHHHHHHHHTTTTTTTTCCCHHHHHHHHHHHHCCCCG-GGCCCCHHHHH

SEQ      LLKDFLRNLKEPLLTFRNRAFMEAAEITDEDNSIAAMYQAVGELPQANRDTLAFIMHL
SEG      .....
COILS    .....
1rgp-    HHHHHHHHTTTTTTGGGHHHHHTTTT-CGGGHHHHHHHHHHHCCCHHHHHHHHHHHH

SEQ      QRVAQSPTKMDVANLAKVFGPTIVAHAVPNPDVMTLQDIKQPKVVERLLSLFLEYWS
SEG      .....
COILS    .....
1rgp-    HHHHHHHHCCCHHHHHHHGGGCC.....

SEQ      QFMVVEQENIDPLHVIENSNAFSTPQTPDIKVSLLGPVTTPEHQLLKTPESSSSLSQRVRS
SEG      .....
COILS    .....
1rgp-    .....

SEQ      TLTKNTPRFGSKSKSATNLGRQGNFFASPLK
SEG      xxx.....
COILS    .....
1rgp-    .....

```

## Prosite for DKFzptes3\_lcl.3

PS00001	212->216	ASN_GLYCOSYLATION	PDOC00001
PS00004	141->145	CAMP_PHOSPHO_SITE	PDOC00004
PS00004	182->186	CAMP_PHOSPHO_SITE	PDOC00004
PS00004	246->250	CAMP_PHOSPHO_SITE	PDOC00004
PS00005	63->66	PKC_PHOSPHO_SITE	PDOC00005
PS00005	174->177	PKC_PHOSPHO_SITE	PDOC00005
PS00005	186->189	PKC_PHOSPHO_SITE	PDOC00005
PS00005	245->248	PKC_PHOSPHO_SITE	PDOC00005
PS00005	313->316	PKC_PHOSPHO_SITE	PDOC00005
PS00005	392->395	PKC_PHOSPHO_SITE	PDOC00005
PS00005	435->438	PKC_PHOSPHO_SITE	PDOC00005
PS00005	595->598	PKC_PHOSPHO_SITE	PDOC00005
PS00005	606->609	PKC_PHOSPHO_SITE	PDOC00005
PS00006	47->51	CK2_PHOSPHO_SITE	PDOC00006
PS00006	66->70	CK2_PHOSPHO_SITE	PDOC00006
PS00006	144->148	CK2_PHOSPHO_SITE	PDOC00006
PS00006	206->210	CK2_PHOSPHO_SITE	PDOC00006
PS00006	234->238	CK2_PHOSPHO_SITE	PDOC00006
PS00006	270->274	CK2_PHOSPHO_SITE	PDOC00006
PS00006	323->327	CK2_PHOSPHO_SITE	PDOC00006
PS00006	387->391	CK2_PHOSPHO_SITE	PDOC00006
PS00006	392->396	CK2_PHOSPHO_SITE	PDOC00006
PS00006	410->414	CK2_PHOSPHO_SITE	PDOC00006
PS00006	449->453	CK2_PHOSPHO_SITE	PDOC00006
PS00006	489->493	CK2_PHOSPHO_SITE	PDOC00006
PS00006	579->583	CK2_PHOSPHO_SITE	PDOC00006
PS00007	46->55	TYR_PHOSPHO_SITE	PDOC00007
PS00007	376->385	TYR_PHOSPHO_SITE	PDOC00007
PS00008	131->137	MYRISTYL	PDOC00008
PS00008	150->156	MYRISTYL	PDOC00008
PS00008	276->282	MYRISTYL	PDOC00008
PS00008	377->383	MYRISTYL	PDOC00008
PS00008	388->394	MYRISTYL	PDOC00008
PS00008	623->629	MYRISTYL	PDOC00008
PS00009	303->307	AMIDATION	PDOC00009

WO 01/12659

PCT/IB00/01496

PS00479 287->336 DAG\_PE\_BINDING\_DOMAIN PDOC00379

Pfam for DKF2phtes3\_lcl.3

HMM_NAME	Phorbol esters / diacylglycerol binding domain		
HMM	*HrEmrHTFrqPTWCDHCgeFIWGwgKQGYOCQnCgMNCHKRCHelVPmm		
	H+F* +T + P +C CG +I +GK ++C +C+++ H +C+ + P		
Query	287	HDFVSKTVIKPESCVPCGKRI-KFGKLSLKRDCRVVSHPECRDRCLP	334
HMM		C*	
		C	
Query	335	C 335	

DKFZphtes3\_lgl3

group: intracellular transport and trafficking

DKFZp DKFZphtes3\_lgl3 encodes a novel 1007 amino acid protein with similarity to human 256 kD golgin.

The new protein contains 7 leucine zippers and seems to be involved in protein-protein-interaction in the golgi apparatus. The very similar rat cpl51 shows haploid-specific transcription in mus musculus testis.

The new protein can find application in modulating protein traffic in the golgi apparatus, especially in human haploid germ cells.

similarity to 256 kD golgi, strong similarity to rat "cpl51"

21 exons encoded on AC004682

EST from a testis library, two mouse ESTs of a testis cDNA library, rat cpl51 shows haploid-specific transcription!

testis or haploid-specific transcription

Sequenced by DKFZ

Locus: map="16q22.2"

Insert length: 3405 bp

Poly A stretch at pos. 3394, polyadenylation signal at pos. 3373

```
1 GGGATAGGGG ATGTGGTTTG TTACAAAGGA TGAGTATTT GATAGCTTCT
51 CATTCTCTGA ACTATTCTGC AGGTTTATAA CAAAGCTCAG AAAATACTAA
101 AGGTTAAAGG AGAATTGAGA GCTGCCAAGG AAATGAAAGA TGAGGCGGGG
151 GAGAGAGACA GAGAAGTGAG CAGCCTGAAC AGCAAGCTGT TAAGCCTGCA
201 ACTTGACATC AAGAATCTGC ACGATGTCTG CAAGAGACAG AGGAAGACCT
251 TGCAGGACAA TCAGCTCTGC ATGGAGGAGG CAATGAACAG CAGCCACGAC
301 AAGAAGCAAG CACAGGCATT AGCATTCTGAG GAGTCAGAGG TGGAAATTTGG
351 GTCCAGTAAA CAGTGTCTATC TGAGACAACCT CCAGCAACTG AAGAAAAAAT
401 TGTGGTCTCT TCAACAAGAA CTGGAGTTTC ACACAGAGGA GTTGCAGACT
451 TCTTACTATT CTCTCCGCCA GTATCAGTCC ATCTAGACA AGCAGACTTC
501 CGACCTGGTT CTCTCGACC ATCACTGCAC ACTGAAAGAA GATGAGGTGA
551 TTCTCTATGA GGAGGAAATG GGAATCACA ACGAGAACAC AGGGGAGAAAG
601 CTCCATTGGG CGCAGGAGCA ACTCGCTTGG GCCGGGGACA AGATCGCCTC
651 TCTAGAGAGG AGCTTAAACC TCTACAGGGA TAAATACCAG TCTTCCCTGA
701 GCAACATCGA GTTACTAGAA TGCCAAAGTA AGATGTTGCA GGGGGAACCTC
751 GGCGGGATCA TGGGTCAGGA GCCTGAGAAC AAGGGTGATC ATTCAAAGGT
801 ACGGATATAC ACTTCTCCTT GCATGATTCA AGAGCATCAG GAGACTCAGA
851 AACGACTGTC TGAAGTCTGG CAAAAGGTCT CTCAACAGGA TGATCTCATT
901 CAAGAACTTC GAAATAAGCT GGCCTGCAGT AACGCTTTGG TTCTGGAGCG
951 TGAAAAGGCT TTGATAAAAC TACAAGCCGA TTTTGCTTCC TGTACAGCCA
1001 CCCACAGATA CCTCTCAGC TCCTCAGGAG AGTGGAAGA CATCAAAAAG
1051 ATACTGAAGC ACTTGCAGGA GCAGAAAGAC AGCCAGTGCC TGCATGTGGA
1101 GGAGTACCAG AACCTGGTGA AGGATCTGCG CGTGGAACCTA GAGGCCGTGT
1151 CGGAACAGAA GAGAAACATC ATGAAGGACA TGATGAAGCT GGAGCTGGAC
1201 CTGCACGGAC TGGCGGAGGA GACATCTGCC CACATTGAGA GGAAGGATAA
1251 GGACATCACC ATCCTGCAGT GCCGGCTGCA GGAGCTGCAG CTGGAGTTCA
1301 CCGAGACCCA AAAGCTCACT TTGAAGAAAG ACAAGTTCTC CCAAGAGAAA
1351 GATGAGATGC TGCAAGAGCT GGAGAAGAAA CTGACACAGG TTCAGAACAG
1401 CCTCTGAAA AAGGAGAAGG AGCTGGAGAA GCAGCAGTGC ATGGCCACAG
1451 AACTTGAAAT GACAGTCAAG GAGGCTAAGC AGGACAAGTC CAAGGAGGCG
1501 GAGTGCAAGG CCTGCGAGGC TGAGGTCCAG AAGCTGAAGA ACAGTCTCGA
1551 AGAGGCCAAG CAGCAGGAGA GGCCTGGTGC TCAGCAAGCA GCCCAGTGCA
1601 AAGAAGAGGC TGCAGTGCCA GGCCTGTACC TGGAGGACAC CCAGAGGAAA
1651 CTGCAGAAGG GTCTCTCCTT GGACAAGCAG AAGGCAGACA CCATCCAGGA
1701 ACTACAGAGA GAACCTCAGA TGCTGCAGAA GGAGTCTCTG ATGGCTGAGA
1751 AGGAACAAAC CTCCAACAGA AAACGGGTGG AGGAGCTGTC ATTAGAACTC
1801 TCTGAAGCCC TGAGGAAGCT TGAATAATCA GACAAGGAAA AGAGGCAGCT
1851 TCAGAAGACA GTGGCTGAGC AGGATATGAA AATGAATGAC ATGCTTGATC
1901 GTATCAAGCA CCAGCACAGG GAGCAAGGCT CCATCAAAAT CAAGTTAGAA
1951 GAAGATCTTC AGGAGGCCAC AAAGCTTCTG GAGGACAAAC GGGAGCAGTT
2001 GAAGAGAGC AAAGAGCATG AGAAGCTGAT GGAGGGAGAA CTTGAAGCTT
2051 TGGCGCAGGA ATTTAAAAAG AAAGACAAGA CTTTGAAGA GAATTCAGA
2101 AAGTTGGAGG AAGAAAAATG GAATCTCCGA GCAGAGCTAC AGTGTGTTT
2151 TACACAACCTG GAATCCTCTC TCACAAATA CAACCCAGC CAGCAAGTCA
2201 TCCAAGACTT GAATAAAGAG ATAGCCCTTC AGAAGGAGTC CTTAATCAGC
2251 CTGAGGCCCT AGCTGGACAA ACCTCTGCAG AAGGAGAGC ACTATCTCCA
2301 GACTACATCC ACCAAAGAAAG CCTATGATGC ATTATCCCGG AAGTCAGCCG
2351 CCTGCCAGGA TCACCTGACA CAAGCCCTCG AGAAGCTCAA TCAGTGACCC
2401 TCAGAGACAA AGAGCTGCA GCAAGCTTG ACACAGACCC AAGAGAAGAA
```

```

2451 AGCTCAGCTG GAAGAGGAAA TCATTGCTTA TGAGGAAAGG ATGAAAAAGC
2501 TCAATACGGA ATTAAGAAAA CTGCGGGGCT TCCACCAGGA GAGTGAGCTG
2551 GAGGTGCAGC CCTTTGACAA GAAGCTAGAG GAGATGAGCT GCCAGGTGCT
2601 GCAGTGGCAG AAGCAACACC AGAATGACCT CAAGATGCTG GCAGCCAAAG
2651 AGGAGCAGCT CAGGGAGTTC CAGGAGGAGA TGGCCGCTT AAAAGAGAAC
2701 CTCCTTGAGG ACGATAAGGA GCCCTGCTGC CTGCCCCAGT GGTCTGTGCC
2751 CAAAGACACC TGTAGGCTCT ACCGAGGGA TGATCAGATT ATGACCAACT
2801 TGGAGCAATG GGCAAAACAG CAGAAGGTGC CCAATGAGAA ACTAGGAAAC
2851 CAGCTCCGAG AGCAGGTGAA CTACATTGCC AAGCTGAGT GCGAAAAGGA
2901 CCACCTCCAC AGTGTAAATG TCCACTTGCA GCAGGAAAC AAGAGCTGA
2951 AGAAGGAGAT AGAAGAGAAG AAGATGAAG CCGAGAACAC AAGGCTATGC
3001 ACCAAGGCC TAGGCCGAG CAGAACGGAG TCCACACAGA GGGAGAAAGT
3051 GTGCGGCACC TTGGGCTGGA AGGGGTGCC CAGGATATG GTTCAAGAA
3101 TGGACCTCAC CAAGTACATC GGGATGCCCC ACTGCCCGG TTCCTCATAC
3151 TGCTAGAATC CACATCTAGC CCTGAGCAGC ATTGCCACGG GTGTTTCTC
3201 AGAGGACAGT GAGTTCCAG CCTCCCTCT CTCTTGACCT GGATCAGCTC
3251 TTACAGGAGT ATATCAGCGT CCCAGCCTAT TTGCAAGAC ACTAACTTT
3301 GTTGAATTT GTCCACTTCC TGCCATGGAG TGAGCTTTAG AACCATACTA
3351 CCATCTCCAG GCCCAAACTC TGAATAAAG ACATGAGCAT GAGCAAAAAA
3401 AAAAA

```

## BLAST Results

Entry AC004682 from database EMBLNEW:  
Homo sapiens Chromosome 16 BAC clone CIT987SK-A-259H10, complete  
sequence.  
Score = 1291, P = 0.0e+00, identities = 265/272

## Medline entries

No Medline entry

## Peptide information for frame 1

ORF from 133 bp to 3153 bp; peptide length: 1007  
Category: similarity to known protein  
Prosite motifs: LEUCINE\_ZIPPER (83-105)  
LEUCINE\_ZIPPER (90-112)  
LEUCINE\_ZIPPER (97-119)  
LEUCINE\_ZIPPER (104-126)  
LEUCINE\_ZIPPER (403-425)  
LEUCINE\_ZIPPER (410-432)  
LEUCINE\_ZIPPER (918-940)

```

1 MKDEAGERDR EVSSLNSKLL SLQLDIKNLH DVCKRQRKTL QDNQLCMEEA
51 MNSSHDKKQA QALAFEESEV EFGSSKQCHL RQLQQLKKKL LVLQOELEFH
101 TEELQTSYYS LRQYSILEK QTSDLVLLHH HCKLKEDEVI LYEEEMGNHN
151 ENTGEKLHLA QEQLALAGDK IASLERSLNL YRDKYQSSLS NIELLECQVK
201 MLOGELGGIM GOEPENKGDH SKVRIYTPC MIQEHQETQK RLSEVWQKVS
251 QODDLIQLR NKLACSNALV LEREKALIKL QADFASCTAT HRYPPSSSEE
301 CEDIKKILKH LQEQKDSQCL HVEEYQNLVK DLRVELEAVS EQRRNHNKDM
351 MKLELDLHGL REETSAHIER KDKDITILQC RLQELQLEFT ETQKLTLLKD
401 KFLQEKOEML QELEKKLTQV QNSLLKKEKE LEKQQCHATE LEHTVKEAQ
451 DKSKEAECKA LQAEVQKLKN SLEBAKQQR LAAQQAQCK EEAALAGCHL
501 EDTQRKLQKG LLLDKQKADT IQELQRELQW LQKESMAEK EOTSNRKRVE
551 ELSLESEAL RKLNSKDEK RQLKTVAEQ DMKMDMLDR IKHQHREQGS
601 IKCKLEEDLQ EATKLEEDKR EQLKKSKEHE KMEGELEAL RQEFKKDKT
651 LKNSRKLEE ENENLRAELQ CCSTQLESSL NKYNTSQOVI QDLNKEIALQ
701 KESLSLQAO LKALQKEKH YLQTTITKEA YDALSRSAA QODDLTQALE
751 KLNHVTSKTH SLOQSLTOTO EKKAQLEEEI IAYEERMKKL NTELRKLRGF
801 HQSELEVHA FDKLEEMSC QVLQWQKHQ NDLKMLAAKE EQLREFQDEM
851 AALKENLLED DKEPCCLPQW SVPKDTCRLY RGNDQIMTNL EOWAKQKVA
901 NEKLGNLRE QVNYIAKLSG EKHDLHSMVY HLQENKKLK KEIEEKKMKA
951 ENTRLCTKAL GPSRTESTQR EKVCGLTGMW GLPQDMGQRM DLTXYIGMPH
1001 CPGSSYC

```

## BLASTP hits

Entry H5417401\_1 from database TREMBL:  
product: "trans-Golgi p230"; Human trans-Golgi p230 mRNA, complete



cds.  
Score = 411, P = 3.9e-34, identities = 212/862, positives = 420/862  
Entry SCINTANA\_1 from database TREMBL:  
Saccharomyces cerevisiae integrin analogue gene, complete cds.  
Score = 404, P = 6.2e-34, identities = 199/897, positives = 423/897  
Entry HS6802\_2 from database TREMBL:  
gene: "MYH9"; product: "dJ6802.2"; Homo sapiens DNA sequence from PAC  
6802 on chromosome 22. Contains apolipoprotein L, myosin heavy chain,  
ESTs, CA repeat, STS and GSS.  
Score = 404, P = 1.9e-33, identities = 231/1028, positives = 469/1028  
Entry AF092090\_1 from database TREMBL:  
product: "cp151"; Rattus norvegicus cp151 mRNA, partial cds.  
Score = 2523, P = 3.0e-262, identities = 506/733, positives = 611/733

## Alert BLASTP hits for DKFZphtes3\_lg13, frame 1

TREMBL:HSGOLGIN\_1 product: "256 kD golgin"; H.sapiens mRNA for golgin,  
N = 1, Score = 411, P = 4.4e-34

TREMBL:HS417401\_1 product: "trans-Golgi p230"; Human trans-Golgi p230  
mRNA, complete Cds., N = 1, Score = 411, P = 4.5e-34

TREMBL:SCINTANA\_1 Saccharomyces cerevisiae integrin analogue gene,  
complete cds., N = 1, Score = 404, P = 7.1e-34

>TREMBL:HSGOLGIN\_1 product: "256 kD golgin"; H.sapiens mRNA for golgin  
Length = 2,185

## HSPs:

Score = 411 (61.7 bits), Expect = 4.4e-34, P = 4.4e-34  
Identities = 212/816 (25%), Positives = 420/816 (51%)

Query: 145 EMGNHNEN-TGEKHLAQEQALAGDKIASLERSLNLYRDYQSSLSNIELLECQVKMLQ 203  
+M + E+ G L +EQL ++ +ERSL+ YR KY ++ ++L+ + K LQ  
Sbjct: 119 DMDSEADLVGNSDSLNKEQLI---QRLRRMERSLSSYRGKYSLVYAYQMLQREKKKQ 175  
Query: 204 GELGGINGQEPENKGDHSEKVIYTSPCMIQEOHQETQKRLSEVMQ-KVSQQDDLIQELRNK 262  
G I+ Q D S RI +Q Q+ +K L E + ++D I L+ +  
Sbjct: 176 G-----ILSQSQ----DKSLRRIAELREELQMDQAKKHLQEEFDASLEEKDQYISVLQ 227  
Query: 263 LAC-----SNALVLEREKALIKLQADFASCTATHRYPPSSSEEC-ED--IKKILKHLQE 313  
++ + + + K L +L+ A P S E ED K L+ LQ+  
Sbjct: 228 VSLKQRLRNGPMNVQVLPQLPQLEPQ-AEVFTKEENPESDGEPPVEDGTSVKTLETLOQ 286  
Query: 314 QKDSQ-----CLH-VEEYQNLVKDLRVELEAVSEQRRNIMKMDMKLELDLHGLREETS 366  
+ Q C ++ ++ L E EA+ EQ ++++ K++ DLH + E+T  
Sbjct: 287 RVKRQENLLKRCETIQSHKEQCTLLTSEKEALQEQDLERLQELERIK-DLH-MAEKTKL 344  
Query: 367 HIERKDKDITILQCRQLQELQLEFFTETQKLTLLKKDKFLQEKDEMLQLEKKLTQV--QNSL 424  
+ +D I Q Q+ + ET++ + + L+ K+E + +L ++ Q+ Q  
Sbjct: 345 ITQLRDAKNLIEQLE-QDKGMVIAETKR---QMHTLEMKEEIAQLRSRIKQMTTGEE 400  
Query: 425 LKKEKELEKQQCMATELEMTVKEAKQDKSKEADCKALQAEVQKLKNSLEEAKQOERLAAQ 484  
L+++KE + + ELE + A+ K++EA K L+AE+ + ++E+ ++ER++ Q  
Sbjct: 401 LREQKE-KSERAAFEELKALSTAQ--KTEARRK-LKAEMDEQIKTIKTSEEERISLQ 456  
Query: 485 QA-AQCKEEAA-LAGCHLEDTQKQLQKGLLLDKQKADTIQELQRELQMLQKSSMAEKEQ 542  
Q ++ K+E + E+ KLQK L +K+ A QEL ++LQ ++E E+ +  
Sbjct: 457 QELSRVQEVVDVMKKSSEEQIAKLQK--LHEKELARKEQELTKKLQTRERE--FOEQMK 512  
Query: 543 TSNRKRVEELSLELSEALRKLENSDKKRLQOKT--VAEQDMKMDMLDRIKHQHQREQGS 600  
+ K E L++S+ + E+ E+ +LQK + E + K+ D+ +  
Sbjct: 513 VALEKSQSEY-LKISQKEQEQESLAELELQKKAILTESENKLRLQEAETVTRILE 571  
Query: 601 IKCKLEEDLQEA TKLLED----KREQLKKSKEHEKLMEG--EELALR-QEFKKKDKTL 651  
++ LE+ LQE +D + E+ K +KE ++E ELE+L+ Q+ + L  
Sbjct: 572 LESSLEKSLQENKNQSKDLAVHLEAKNKHKNKEITVMVEKHKTELESKHKQDQALWTEKL 631  
Query: 652 KENSRLKEEENENLRAELQCCSTQLESSL-NKYNTSQQVIQDLNKE----IALQKESLMS 706  
+ ++ + E E LR + C + E+ L +K Q I+++N++ + +++ L S  
Sbjct: 632 QVLKQQYQTEMKLRK---CEQEKETLLKDKKEIIFQAHIEEMNKTLEKLDVQKQTELES 688  
Query: 707 LQAQLDKALQKEKHLYLT--TITKEAYDALSRKSAACQDDLTQALEKLNHVTSKSLQ 764  
L ++L + L K +H L+ ++K+ D + ++ A D+ Q V S K +

Sbjct: 689 LSSELEVL-KARHKEEELSVLKDQTKMKQLEAKMDE--QKNHHQQQVDSIIKEHEV 745

Query: 765 SLTQTQEKKAQLEEEIIAYEERMKKLNTLRLRGFHESELEVHAFDKKLEEMSCQVLQ 824  
S+ +T+ KA L+++I E +K+ + L++ + + E++ + +L++ S++

Sbjct: 746 SIQRTE--KA-LKDQINQLELLKKERDKHLKEHQAHVENLEADIKRSEGLQASAKLDV 802

Query: 825 WQKQHNDLKLAAKEEQLEFQEEAALKENLLEDDKEPCCLPQW-----SVPKDT-C-R 878  
+Q +Q+ A EQ + ++E++A L++ LL+ + E L + + KD C

Sbjct: 803 FQS-YQS-----ATHEQTKAYEEQLAQLOQLLDOLETERILTRQVAEVAQKKDVCTE 855

Query: 879 LYRGNDQIMTNLEQWAKQKQVANEKLGNLREQVNYIAKLS-GEKDLHLSVMVHLQEQNK 937  
L Q+ ++Q RQ +K+ + QV Y +KL G K+ + + ++EN

Sbjct: 856 LDAHKIQVDLMQQLKQNSEMEQKVSILT--QV-YESKLEDGNKEQEQTKQILVEKENM 912

Query: 938 KLK-KEIEEKKMAENTRLCTK 958  
L+ +E ++K+++ +L K

Sbjct: 913 ILQMRGQKKEIEILTKLSAK 934

Score = 338 (50.7 bits), Expect = 3.1e-26, P = 3.1e-26  
Identities = 216/953 (22%), Positives = 468/953 (49%)

Query: 2 KDEAGERDRE--VSSLNS-KLL-SLQLDIKNLHDVCKRQKTLQDN-QLCM-----EEM 51  
K+E E D E V S K L +LQ +K ++ RR ++T+Q + + C +EA+

Sbjct: 260 KEENPESDGEPPVEDGTSVKTLETQQRVQRQENLLRCKETIQSHKEQCTLLTSEKEAL 319

Query: 52 NSSDKKQAQALAFESEVEFGSSKCHLRQ---LQOLK--KKLLVLQLEFHTTEELQ 105  
D++ + ++ + + LR ++QL+ K +++ + + H E L+

Sbjct: 320 QEQLDERLQLEKIKIDLHMAEKTKLITQLRDAKNLIEQLEQDKGMVIAETKRQMH-ETLE 378

Query: 106 TSYSYLRQYSILEKQTSDDLVLHHCKLKEDEVILYEEEMGNHNTENGKHLHQAQQL- 164  
+ Q +S +++ T+ L K K + E E +T+K A+ +L

Sbjct: 379 MKEEIAQLRSRIKQMTTQGEELREQ-KEKSERAAFEELKAL--STAQTEEARRLK 434

Query: 165 ALAGDKIASLERSLNLRYDKYQSSLSNI--ELLECVKMLQELGGIMGOEPENKGDHKS 222  
A ++I ++E++ R Q LS + E+++ K + ++ + Q+ K K

Sbjct: 435 AEMDEQIKTIKTSEERISLQQLSRVKQEVVDVMKSSSEEQIAKL--QKLHEKELARK 492

Query: 223 VRIYSPCMIQEHQETQKRLSEVMQKVSQDDLIQELRNKLACSNALVLEREKALIKLOA 282  
+ T +E +E Q+++ +K SQ + L ++ + +L LE ++LQ

Sbjct: 493 EQELTKKLQRE-REFEQMKVALEK-SQSEYL--KISQKEQESLAE--LELQK 544

Query: 283 DFASCTATHRYPPSSSECEDIKKILKHLQEKDQSCQLHVEEYQNLVLDLVELEAV-SE 341  
A T + +E E + + L+ + + ++E +N KDL V LEA ++

Sbjct: 545 K-AILTESENKLRLDQEAETRYTRILELESSLEKS---LQENKQSKDLAVHLEAEKNK 600

Query: 342 QRRNIMKMKLELDLHGLREETSAMERKDKDITI-LQCRQLQLEQLEFTETQKTLTKKD 400  
+ I + K + +L L+ + A K + + Q +++L+ E +K TL KD

Sbjct: 601 HNKEITVMVERKHTELESKHHQDALWTEKLQVLKQYQYTEMKLR-EKCEQEKETLLKD 659

Query: 401 K-----FLQEKDEM-LQLEKKLTQVQNSLLKKEKELEKQCMATELEMTVKEAQDKS 453  
K ++E +E L++L+ K T+++ SL + E+ K + E +E+V + + DK

Sbjct: 660 KEIFQAHIEEMNEKTEKLDVKQTELE-SLSELESEVLKARHKEE-ELSVLKDQTKM 717

Query: 454 K-EAECKALQAEVQKLKNSLEEAQQRRLAAQQAQC-KEEAALAGCHLEDQTKLQKGL 511  
K E E K + + ++ ++ ++ Q+ + K++ L++ + L++

Sbjct: 718 KQLEAK-MDEQKNHHQQQVDSIIKEHEVSIQRTEKALKDQINQLELLKKERDKHLKEHQ 776

Query: 512 L-LDKQKADTIQELQRELQMLQKESMAEKEQTSNRKRVEELSLELSEALRKLNSDKEK 570  
++ +AD I+ + ELQ + + + Q++ ++ + +L++ +KL + + E+

Sbjct: 777 AHVENLEAD-IKRSEGLQASAKLDVFSYQSATHEQTKAYEEQLAQLOQLLDOLETER 835

Query: 571 RQLQKTVAEQDMKMDM---LD--RIKHQHQSGSIK--CKLEEDLQATKLLDKREQL 623  
L K VAE + + D+ LD +I+ Q Q K ++E+ ++ T++ E K E

Sbjct: 836 ILLTKQVAEVAQKKDVCTELDAHKIQVDLMQQLKQNSEMEQKVSILTQVYESKLEDG 895

Query: 624 KKSKEHEK--LMEGELEALRQEFKKDKTLKENSRLKEENENLRAELQCCSTOLESSLN 681  
K +E K L+E E L+ +K K ++ ++KL + +++ + T+ ++

Sbjct: 896 NKEQEQTKQILVEKENMILQMRGQK-KEIEILTQKLSAKEDSIHILNEEYETKFNQEK 954

Query: 682 KYNTSQOVIQDLNKEIALQKESLMSLQAQLDKALQEKHYLQTTITKEAYDALSRKSAAC 741  
K Q +++ + + K+ L+ +A+L K L E L+ + ++ ++A + A

Sbjct: 955 KMEKVKQAKEMOETL---KKLLDQEAKLKEL--ENTALELSQREKQFNAMLEMAQA 1009

Query: 742 QD-DLQALEKLNHVTSKSLQOSLTQTQEKKAQLEEEIIAYEERMKKLNTLRLRGF 800  
++ A+ +L T++ + ++ SLT+ + +L + I +E KKLN + +L+

Sbjct: 1010 NSAGISDAVSRLE--TNQKEQIE-SLTVHRR--ELNDVISIWE---KKLNQAEELQEI 1061

Query: 801 HQSELEVHAFDKKLEEMSCQVLQW--QKQHNDLKLAAKEEQLEFQEEAALKENLL 858  
H E+++ ++++ E+ ++L + +K+ N ++ KEE +++ + L+E L

Sbjct: 1062 H---EIQLQEKQEVAEKQKILLFGCEKEEMNK-EITWLKEEGVKQ-DTTLNELQEQKL 1116

Query: 859 EDDKEPCCLPQWVSPKDTCLRYRGNDQIMTNLEQ--WAKQKQVANEKLGQNLREQVNYI- 915  
+ L Q K L + + +L++ + ++Q V + L + + +V+ +  
Sbjct: 1117 QKSAHVNSLAQ-DETKLKAHLEKLEVDLNKSLKENTFLOEQVLKMLAEEDKRRKVELT 1175

Query: 916 AKLSGEKDLHLSVMVHLQEQENKKL-KEIEEKKMAE 951  
+KL + S+ ++ NK L+ K +E KK+ E  
Sbjct: 1176 SKLKTDEEFQSLKSSHEKSNKSLKEDKSLKLEFKKLE 1212

Score = 337 (50.6 bits), Expect = 4.0e-26, P = 4.0e-26  
Identities = 215/951 (22%), Positives = 433/951 (45%)

Query: 10 REVSSLNSKLLSLQDLKLNLDVCKRQKTLQDNQLCMEAMNSSHDKKQAALAFEESE 69  
+E + +++L L+ ++ K Q K L + EA + H+K+ + E+ +  
Sbjct: 560 QEAETRYTRILELESSLEKSLQENKQSKDLAVHL---EAEKNKHNKEIT--VMVEKHK 613

Query: 70 VEFSSKQCHLRQLQQLKKLLVLQOELEFHEELOTSSYSLRQYQSILEKQTSDLVLLH 129  
E S K H +Q +KL VL+Q+ + E+L+ Q + L K +++  
Sbjct: 614 TELESK--H-QQDALWTEKLQVLKQYQTEMEKREK---CEQKETLLKD-KEIIFQA 666

Query: 130 HHCKLKE---DEVILYEEEMGNHENTGKEL---HLAQQLALAGDKIASLERSLNLRYD 183  
H +E +++ +E+ + + E L H +E+L+ D+ +++ L D  
Sbjct: 667 HIEEMNEKTLEKLDVKQTELESLSSESEVLKARHKEEELSVLKQDTDKMKELEAKMD 726

Query: 184 K---YQSSLSNIELLECQVKMLQGE--LGGIMQEPENKGOHKSIVRIYTPCMIEHQE 237  
+ +Q + +I + E +V + + E L + Q + K + + + +  
Sbjct: 727 EQKNHQQQVDSI-IKEHEVSIQRTKALKDQINQLELLKRDK-HLKEHQAHVENLEA 784

Query: 238 TOKRLSEVMQKVSQDDLIQELRNKLACSNALVLEREKALIKLQADFASCTATHRYPPSS 297  
KR Q+ S + D+ Q ++ ++ E+ L+Q T R  
Sbjct: 785 DIKRSEGLQASAKLDVFSYQS---ATHEQTKAYEEQLAQQLKLDLE-TERIL--- 837

Query: 298 SEECEDIKKILKHLQEQKDSQCLHVEEYQNLVKDLRVELEAVSEQRNIMKMMKL-ELD 356  
+ K + ++ QK C ++ ++ V+DL +LE + + +K + ++ E  
Sbjct: 838 -----LTKQVAEVEAQKDVCTELDAHKIQVQDLMOQLEKQNSEMEQVKSITQVYESK 891

Query: 357 LH-GLREETSABIERKDKDITILQCLRL-QELQLEFTEQKTLKDKDF--LQEKDEM-LQ 411  
L G +E+ +K+ ILQ R Q+ ++E TOKL+ K+D L E+ E +  
Sbjct: 892 LEDGNKEQEQTQKQILVEKENMILQMRQKKEIIL-TQKLSAKEDSIHILNEEYETKFK 950

Query: 412 ELEKILTVQNSLLK-----KEKELEKQCCMATELEMTVKEAKQDKSKEAECKALQAEVQ 466  
EKK+ +V+ + K+K L+++ + ELE T E Q K K+ K L+ Q  
Sbjct: 951 NQEKMEKVKQKAKEMQETLKKLLDQEAELKKELENTALELSQ-KEKQFNAMKLEM-AQ 1008

Query: 467 KLNLSLEEAQQLERLAAQQAQCKEEAALAGCHLEPTQRKLQKGLLLDKQADTIQELQR 526  
+ +A RL Q Q + + L D +K L Q+A+ +QE+  
Sbjct: 1009 ANSAGISDAVS--RLETNQEQIESLTVHRELDNVISIWEKKL---NQAEELQEIH- 1062

Query: 527 ELQMLQESSMAEKEQT-----SNRKR--EELSLESEALRKLNSDKKRLQ 574  
E+Q+ +KE +AE +Q K + +E ++ L +L+ K+K  
Sbjct: 1063 EIQLQEKEQVAVELKQKILLFGCEKEEMKEITWLKEGVKQDTTLNELQELKQKSAHV 1122

Query: 575 KTVAEQDMKMDMLDRIKHOREQGSIRCKLEEDLQEA TKLLEDKREQLKKSKEHEKLME 634  
++A+ + K+ L+++ + L+E L E L E+ + ++ + K +  
Sbjct: 1123 NSLAQDETCLKAHLEKLEVDLNKSLKENTFLOEQVLKMLAEEDKRRKVELTSLKLTDD 1182

Query: 635 GELEALRQEFKKKDKRTLKENSRLKEEENENLRAELQCCSTQLESSLNKYNTSQQVIQDLN 694  
E ++L+ +K +K+L++ S + ++ +E L +L C + E+ L T++ + +  
Sbjct: 1183 EEFQSLKSSHEKSNKSLKEDKSLKLEFKLSEELAIQDLICCKKTEALLEA-KTNELINISS 1241

Query: 695 KEIALQKESLMSLQAQLDKALQKEKHLYLTTITKEAYDALSRKSAACQDDLT---QALE 750  
K A+ + Q + K KE ++T E +A R+ Q+ L QA  
Sbjct: 1242 KTNAILSR-ISHCQHRRTKV--KEALLIKTCTVSEL-EAQLRQLTEEQNTLNISFQQATH 1297

Query: 751 KLNHVTSETKSLQSLTQTEKKAQLEEEIAYEERMKKLN---TELK--LRGFHQESE 805  
+L ++ KS++ + +K L++E ++ + T+L+K + +  
Sbjct: 1298 QLEEKENQIKSMKADIESLVTEKEALQKEGGNQQAASEKESCITQLKELSENINAVTL 1357

Query: 806 LEVHAFDKKLE--EMSCQVLQWQKHONDKMLAAKEEQLEFQEEAALKENLLEDDKE 863  
++ +KK+E +S Q+ Q QN + L+ KE + +++ K LL D +  
Sbjct: 1358 MKEELKKEKKEISSLSKQLTDLNVQLQNSIS-LSEKAAISSLRKQYDEEKCELL-DQVQ 1415

Query: 864 PCLLPQWVSPKDTCLRYRGNDQIMTNLEQWAKQKQVANEKLGQNLRE---QVNYIAKLSG 920  
++ K+ D +W K+ + + N ++E Q+ +K +  
Sbjct: 1416 DLSFRVDTLSKEKISALEQVDDWSNKFSEMKKKAQSRFTQHONTVKELQIQLELKSKEY 1475

Query: 921 EKDH-LHSVMMVHLQEQENKK---LKKEIEEKKMAE 951  
EKD ++ + L Q+NK+ LK E+E+ K K E  
Sbjct: 1476 EKDEQINLLKEELDQONKRFDCLEKGMEDDKSKME 1510

Score = 332 (49.8 bits), Expect = 1.4e-25, P = 1.4e-25  
Identities = 209/953 (21%), Positives = 438/953 (45%)

Query: 1 MKDEAGERDREVSSLSKLLSLQLDIKNLHDVCKRQKRTLDNQLCMEAMNS---SHD 56  
 Sbjct: 470 MKKSSEQIAKLQKLHEKELARK-EQELTKKLTREREFEQMKVALEKSQSEYLKISQE 528

Query: 57 KKQAQALAFEESEVEFGSSKQCHLRQLQQLKKLLVLQOELEFTEELQTSYSLRQYQS 116  
 Sbjct: 529 KEQQESLALAELELQ---KKAILTESEN---KLRLQQAETRYRTRILESSLEKSLQ 581

Query: 117 ILEKQTSDLVLLHHHCKLKEDE--VILYEE----EMGNHNENT--GEKLHLAQEQALALA 167  
 Sbjct: 582 ENKNQSKDLAVHLEAEKNKHNKEITVMVEKHKTELESLEKHQDQDALWTEKLQVLKQYQTE 641

Query: 168 GDKIASL--ERSLNLYRDK---YQSSLS--NIELEECQVKMLQGEELGGIMGQEPENKGDH 220  
 Sbjct: 642 MEKLREKCEQEKETLLDKKEIIFQAHIEEMNEKTLE-KLDVKQTELESLSSELSEVLKAR 700

Query: 221 SKVRIYTPCMIQEHQETQKRLSEVWQKVSQDDLIQELRNKLACSNALVLEREKALIKL 280  
 Sbjct: 701 HKLEELS--VLKD--QTDKMKQLEAKMDEQKNHQQQVDSIIKEHEVSIQRTKALKD 756

Query: 281 QADFASCTATHR--YPPSSSEECEDIKKILKHLQEQKDSQCLHVEEYQNLVKDLRVELEA 338  
 Sbjct: 757 QINQLELLKERDKHLKEHQAHVENLEADIKRSEGEQQASAKLDVFSQYSQATHEQTKA 816

Query: 339 VSEQKRNMKDMKLELDLHGLREETSABIERKOKDITILQCRLOELQLETFETQKLT 398  
 Sbjct: 817 YEEQLAQQLKLLDLETERILLTKQV-AEVEAQKQDV---CT--ELDAHKIQVQDLMOQ 869

Query: 399 KDKFLQEKDEMLQLEKLLTQVNSLLKK-EKELEKQOCMATELEMTVKEAKQDKSKEAE 457  
 Sbjct: 870 LEK---QNSEMEQKV-KSLTQVYESKLEDGKKEQEQTKILVEKENMILQMGREGQKKEIE 925

Query: 458 C--KALQAEVQKLKNSLEEAKQOERLAAQAAQCKEEAALAGCHLEDQTK--LQKGLLL 513  
 Sbjct: 926 ILTQKLSAKEDSIHILNEEYETKFNQKKEKMEKVKQKAK---EMQETLKKKLLDQEA 981

Query: 514 DKQKADTIQEL-QRELQMLQKESMAEKQTSNRKRVEELSLELSEALRKLNSDKERKQ 572  
 Sbjct: 982 KKELENTALELSQKQEKFNQAKMLEMAQANSAGISDAVSRLETNQKEQIESL--TEVHRRE 1039

Query: 573 LQKTVAEQDKMMDMLDRIKHQHREQGSIKCKLEEDLQEA TKLEEDKREQLKKS---KE 628  
 Sbjct: 1040 LNDVISIWEKKLNQQAEELEQIHEIQLEKEQEVAELKQKILLFGCEKEEMNKETWLKE 1099

Query: 629 HEKLMEGEALRQEFKKKDKTLKENSRLKEENENLRAELQCCSTOLESSLNKYNTSQ 688  
 Sbjct: 1100 EGVKQDTTLNELQEQKSAHV--NS--LAQDETKLKAHLEKLEVDLNSLKTENTFLQE 1155

Query: 689 VIQDLNKEIALQKESLMSLQAQL---DKALQ--KEKHYLQTTITKEA---YDALSRSKAA 740  
 Sbjct: 1156 QLVELKMLAEEDKRKVSELTSKLTDEEFQSLKSSHEKSNKSLDKSLEFKKLSSE-LA 1214

Query: 741 QDDDL----TQAL-----EKLNHVTSETKSLQSLTQTOEKKAQLEEEI IAYEERMKKL 790  
 Sbjct: 1215 IQLDICCKTEALLEAKTNELINISSKTNAILSRISHCQHRRTTKVKEALLIKTCTVSEL 1274

Query: 791 NTELRLRGFQHESELEVHAFDKLEEMSCQVLQWQKHQNDLKMMLAAKEEQLEFQEM 850  
 Sbjct: 1275 EAQLRQLTEEQNTLNISFQQATHOLEEKENQI---KSMKADIESLVTEKEAL---QKEG 1327

Query: 851 AALKENLLEDDKEPCCLPQWVSPKDTCLRYRGNDQIMTNLEQWAKQKQVANEKLGNOLE 910  
 Sbjct: 1328 G--NQQAASEKESC-ITQ--LKKELSE---NINAVTLMKEELKEKKVEISSLSKQLTD 1378

Query: 911 ---QVNYIAKLSGEKDLHLSVMVHLQENKKLKEIEEKKMAE 951  
 Sbjct: 1379 LNVQLQNSISLSEKAAISSLRKQYDEEKCELLDQVQDLSFKVD 1422

Score = 329 (49.4 bits), Expect = 2.9e-25, P = 2.9e-25  
 Identities = 226/941 (24%), Positives = 444/941 (47%)

Query: 61 QALAFEESEVE--FGSSKQCHLRQLQQLKKLLVLQOELEFTEELQTSYSLRQYQSIL 118  
 Sbjct: 165 QMLQREKKLQGLSQQSQQSLRRIAELREELQMDQAKHLQEEFDASLEEKQYISVL 224

Query: 119 EKQTSDLVLLHHHCKLKEDEV-----ILYEEEMGNHNENT---GEKL---HLAQEQALALA 167  
 Sbjct: 225 QTVSLLKQRLRNGPMNVVLPQLEPQAEVFTKEENPESDGEPVVEDGTSVKLTETL 284

Query: 168 GDKIASLERSLNLYRDKYQSSLSNIELEECQVKMLQGEELGGIMGQEPENKGDHDKSVRIYT 227

Sbjct: 285 ++ E L ++ QS LL ++ LQ +L + QE E D ++ 340  
 Query: 228 SPCMIQEHQETQKRLSEVWQKVSQDDLIQELRNKLACSNALVLEREKALIKLQADFASC 287  
 +I + ++ + ++ Q +I E ++ L ++ E+ +L++  
 Sbjct: 341 KTKLITQLRDAKNLIEQLEQDKGM--VIAETKRQM--HETLEMKEEE-IAQLRSRIKQM 394  
 Query: 288 TATH--RYPPSSSEEC--EDIKKILKHLQEQKDSQCLHVEEYQNLVKDL-----RVE 335  
 T R SE E+++K L Q+ +++ E +K + R+  
 Sbjct: 395 TTQGEELREQKEKSERAAFEELEKALSTAQKTEARRKLKAEQDEQIKTIEKTSEERIS 454  
 Query: 336 LEA-VSEQKRNIMKMMKL--ELDLHGLREETSABIERKDKDITILQCRLOELQLETTET 392  
 L+ +S K+ ++ D+MK E + L++ + RK++++T +LQ + EF E  
 Sbjct: 455 LQQELSRVKQEVV-DVMKKSSEEQIAKLKHEKELARKEQELTK--KLQTREREFQEQ 510  
 Query: 393 QKLTLLKKDKFLQEKDEMLQLEKLTQVQNSLLKKEKELEKQCCMATELEMTVKEAKQDK 452  
 K+ L+K + E ++ QE E+ Q SL +E EL+K+ + TE E +++ +Q+  
 Sbjct: 511 MKVALESQ--SEYLIKISQEQE-----QESLALZEELELQKKAIL--TESENKLRDLQEQ- 561  
 Query: 453 SKEAECKALQAEVQKLNKNSLEAKQER-----LAAQQAQCKEEAALAGCHLEDTOR-K 506  
 ++ + L+ E L+ SL+E K Q + L A++ KE + H + + K  
 Sbjct: 562 AETYRTRILELE-SSLEKSLQENKQSKDLAVHLEAKNKNKEITVMVEKHKTELESK 620  
 Query: 507 LQGLLLDKQKADTIQELQRELQMLQKESMAEKEQTSNRKRVEELSLELSEALRK-LEN 565  
 Q+ L ++ Q+ Q E++ L +E EKE K + + E K LE  
 Sbjct: 621 HQQDALWTEKLQVLKQYQTEMEKL-REKCEQEKETLLKDKKII-FOAHIEEMNEKTLEK 678  
 Query: 566 SDKEKRLQKTVAEQDMKMNMDLRIKHQHQGSI-KCKLEEDLQEA-TKLEDDR--E 621  
 D ++ +L+ +E +++L + +H+ E+ S+ K + ++ QE K+ E K +  
 Sbjct: 679 LDVKQTELESLSSE----LSEVL-KARHKLEELSVLKQDQDKMKQLEAKMDEQKNHQ 733  
 Query: 622 QLKKS--KEHEKLMGELEALRQEFKKDKDTLKENSRLKEEN--ENLRAELQCCSTQL 676  
 Q S KEHE ++ +AL+ + + LKE + L+E ENL A+++ +L  
 Sbjct: 734 QQVDSIIKEHEVSIQRTKALKDQINQLELLKERDKHLKEHQAHVENLEADIKRSEGL 793  
 Query: 677 ESSLNKYNTSQVIQDLNKEIALQKESLMSLQALQALQKQKHLYLQTTITKEAYDALSR 736  
 + + K + Q +++ +E L LQ +L L+ E+ L TK+ + ++  
 Sbjct: 794 QQASAKLDVFSYQSATHEQTKAYEEQLAQLQKRL-LDLETERILL----TKQVAEVEAQ 848  
 Query: 737 KSAACQD-----DLTQALEKLNHVTSETKSLQOSLTQOEKKAQ--LEEEIIAYEE 785  
 K C + DL Q LEK N SE + +SLTQ E K + +E+ +  
 Sbjct: 849 KKDVTCLDAHKIQVQDLMQLEKQ--SEMEQKVSLSLTQVYESKLEDGKKEQEQTKQI 905  
 Query: 786 RMKLNTELRLKRGFHOESELEVHAFDKKLEEMSCVIL--QWQKHQNDLMLAAKEEQ 843  
 ++K N L+ G Q+ E+E+ +L +E S +L +++ +N K + ++  
 Sbjct: 906 LVEKENMILQMRG--QKKEIILTQKLSAKEDSIHILNEEYETKFNQEKKMEKVKQKA 963  
 Query: 844 REFOEMAALKENLLEDDKEPCCLPQWVSKDTCLYRGNDQIMTNLEQWAKQKV---- 899  
 +E QE LK+ LL+ ++ L + + L + Q + + A+  
 Sbjct: 964 KEMQE--TLKKKLLDQEAQ--LKK-ELENTALELSQKEQKFNKMLEMAQANSAGISD 1016  
 Query: 900 ANEKLGNLREQVNYIAKLSG-EKDHLSVMVH-LQENKKLKK--EIEEKKMAENTRL 955  
 A +L +EO+ + ++ E + + S+ L Q+ ++L++ E+ + + E L  
 Sbjct: 1017 AVSRLETNKEQIESLSTEVRRLNOVISIWEKKLNQAEELQEIHEIQLEKEQEVAEL 1076  
 Query: 956 CTKALGPSRTESTQREKVCGLGWKGLPQD 985  
 K L E + K L +G+ QD  
 Sbjct: 1077 KQKIL-LFGCEKEEMNKEITWLKEEGVKQD 1105  
 Score = 326 (48.9 bits), Expect = 6.0e-25, P = 6.0e-25  
 Identities = 220/907 (24%), Positives = 444/907 (48%)  
 Query: 67 ESEVEFGSSKQCHLRQLQQLKKLLVLQLELFHTEELQTSYSLRQYQSIIE---KQTS 123  
 E+E G+S + QL Q +++ EL T+Y L++ + L+ Q+  
 Sbjct: 123 EAEDLVGNSDSLNEQLIQLRRMERSLSSYRGKYSSELVTAYQMLQREKKKLQGISQSQ 182  
 Query: 124 DLVLLHHCKLKEDEVILYEEEMGNHNTENGKHLAQEQALAGDKIASLERSLNLYRD 183  
 D L +L+E+ + +++ H+ E+ +E+ I+ L+ ++L +  
 Sbjct: 183 DKSL-RRIAELREE--LQMDQAKKHLQ--EEFDASLEE---KDQYISVLQTVSLLKQ 233  
 Query: 184 KYQSSLSNIELLECQVKMLQGEELGGIMGOE-PENKG-----DHKVR-ITYSPCMIOEHQ 236  
 + ++ N+++L+ + L+ + +E PE+ G D + V+ + T ++ +  
 Sbjct: 234 RLANGPMNVDLK-PLPQLEPQAEVFTKEENPESDGEFVVEDGTSVKTLETQORVKRQE 292  
 Query: 237 ETQKRLSEVWQKVSQDDLIQELRNKLACSNALVLEREKALIKLQADFASCATATHRYPPS 296  
 ER E Q +Q L+ KA L ER + L K+ + D T  
 Sbjct: 293 NLLKRCKETIQSHKEQCTLLTS--EKEALQEQLD-ERLQELEKIK-DLHMAEKTLLIT-- 346  
 Query: 297 SSEECEDIKKILKHLQEQKDSQCLHVEEYQNLVKDLRVELEAVSEQKRNIMKMMKLELD 356  
 + D K +++ L++ K + E + + + L ++ E ++ Q R+ +K M +  
 Sbjct: 347 ---QLRDAKNLIEQLEQDKGM--VIAETKRQMHEHETLEMKEEEIA-QLRSRIKQMTTQGE 400

Query: 357 LHGLREETS-AHIERKDKDITILQCRLOE----LQLEFTETQKLTLLKKDKFLQEKDEMLQ 411  
L +E++ A E +K ++ Q + +E L+ E E K T++K +E+ + Q  
Sbjct: 401 LREQEKESERAAFELEKALSTAQ-KTEEARRLKAEMDEQIK-TIEKTSE-EERISLQQ 457

Query: 412 ELEKKLTQVQNSLLKK-EKELEKQCCMATELEMTVKEAKQDKSKEAECKALQAEVQKLN 470  
EL + +V + + K E+++ K Q + E E+ KE Q+ +K+ + + + Q +K  
Sbjct: 458 ELSRVKQEVVDVHKSSSEEQIAKLQKLH-EKELARKE--QELTKKLQTREREFQEQ-MKV 513

Query: 471 SLEAAQQERLAAQQAQCKEAAAGCHLEDOTQKRLQ-KGLLLD-KQKADTIQELQREL 528  
+LE++ Q E L Q + +E AL L+ + + L D +Q+A+T + EL  
Sbjct: 514 ALEKS-QSEYLIKISQEKQESLAELELQKKAILTESENKRLQDQEAETRYTRILEL 572

Query: 529 QMLQKESMAEKEQTSNRKRVEELSLELSEALRLKLENS-DKEKRLQKTVAEQDMKMDM 587  
+ E S+ E + S V L E ++ +++ +R K +L+ +QD +  
Sbjct: 573 ES-SLEKSLQENQSKDLAVH-LEAEKNKHKEITVMVEKHKTELESLKHQDALWTEK 630

Query: 588 LDRIKHQHR-EOGSICKLEEDLQEAATKLEDKRE--QLKKSKEHEKLMEGELEALRQEF 644  
L +K Q++ E ++ K E OE LL+OK Q + +EK +E +L+ + E  
Sbjct: 631 LQVLKQQYQTEMEKLEKCE---QEKETLLKDKKEIIQAHIEEMNEKTE-LKDVVKTEL 686

Query: 645 KKKDKTLKE--NSR-KLEENENLRAELQCCSTQLESSLNKY-NTSQVQIDLNKE--IA 698  
+ L E +R KLEEE L+ + +LE+ +++ N QQ + + KE ++  
Sbjct: 687 ESLSELSEVLKARHKLKEELSVLKDOTDKMKQLEAKMDEQKNHQQQVDSIIEKEVS 746

Query: 699 LQK-ESLMSLQA-QLDKAL-QKEKHLYLTITKEAYDALSRKS-----AACQDDLTQAL 749  
+Q+ E + Q QL+ L +++KH + E +A ++S A+ +D+ Q+  
Sbjct: 747 IQRTKALKDQINQLELLKDKLKHLEQAHVENLEADIKRSEGELOQASAKLDVFSY 806

Query: 750 EKLNHVTSETKSLQOSLTQTEKKAQLEEEIIAYEERMKMLTELRLKGFHQESELEVH 809  
+ H +TK+ ++ L Q Q+K LE E I +++ + + + +  
Sbjct: 807 QSATH--EQTKAYEEQLAQQLKLLDLETERILLTKQVAEVAQRKDVCTELDAHKIQV 864

Query: 810 AFDKKEEMSCQVLQWQKHQN--DLKMLAAKEQLREFQEEMAALKENLL----EDDK 863  
++LE+ ++ Q K + K+ +EQ E +++ KEN++ E K+  
Sbjct: 865 DLMQQLKQNSMEQVKSLTQVYESKLEDGNKEQ--EQTKQILVEKENHILQMRGQKK 922

Query: 864 PC-CLPO-WSVPKDTCLRYRGNDQIMTNLE-QWAKQKVANE--KLGNQLREQV-NYIAK 917  
L Q S +D+ + N++ T + Q K +KV + ++ L++++ + AK  
Sbjct: 923 EIEILTQKLSAKEDSIHIL--NEEYETKFNQEKMEKVKQKAKEMQETLKKKLQDEAK 980

Query: 918 LSGEKDHLHSMVHLQQENKLLKKEIEKKMKAEENTRLCTKALGPSRTSTESTOREKV 973  
L K L + + L Q+ K+ ++ E M N+ + A+ SR E+ Q+E++  
Sbjct: 981 L---KKELENTALELSQKEKQFNAKMLE--MAQANSAGISDAV--SRLETNQEQI 1029

Score = 318 (47.7 bits), Expect = 4.4e-24, P = 4.4e-24  
Identities = 184/827 (22%), Positives = 405/827 (48%)

Query: 1 MKDEAGERDREVSSSLNSKLLSLQLDIKNLHDVCKRQRKTLQDNQLCMEAMNSHDKK-Q 59  
++ E G + + S S + L+ ++ + ++ L++ ++ + D Q  
Sbjct: 1323 LQKEGGNQQAASEKESCITQLKKELSENINAVTLMKEELKEKKVEISSLSKQLTDLNVQ 1382

Query: 60 AQ-ALAFESEVEFGSSKQCHLRQLQQLKKLLVLQOELEFHTTELQTSYYS-LRQYQS- 116  
Q +++ E E S + +Q + K +LL Q+L F + L S L Q  
Sbjct: 1383 LQNSISLSEKAAISSLR----KQYDEEKCELQDQVQDLSFKVDLSREKISALEQVDDW 1438

Query: 117 ---ILE-KQTSIDLVLHHCKLKEDEVILYEEEMGNHNTGKELHLAQEQALAGDKIA 172  
E K+ + H +KE ++ L + + ++ E+++L +E+L +  
Sbjct: 1439 SNKFSEWKKKAQSRFTQHQNTVKELQIQLELSKEAYEKD--EQINLLKEELDQNNKRPD 1496

Query: 173 SLERSLNLYRDKYQSSLSNIEL-LECOVKMLQELGGIMQEP-ENKGDHKSIVRIYTSPC 230  
L+ + + K + SN+E L+ Q + EL + Q+ E + + ++ Y  
Sbjct: 1497 CLKGEMEDDKSKMEKESNLETLSQTARIM-ELEDHITQKTIEIESLNEVLKNYNQK 1555

Query: 231 MIQEHQETQKRLSEVWQKVSQDDLIQELRNKLACSNAVLEREKALIKLQADFASCTAT 290  
I EH+E ++L + ++D+ +E K+ L LE + +K + +  
Sbjct: 1556 DI-EHKELVQKLQHFQELGEEDNVRKEAEKI-----LTLENQVYSMAELETKKKELE 1609

Query: 291 HRYPPSSSEECEDIKKILKHLEQKDSQCLHVE-EYQNLVKDLRVELEAVSEQKRNIMKD 349  
H S+E E+K + L+ + + + + + + + +L + E+K ++  
Sbjct: 1610 HVNLSVKSE-EELKALEDRLSESAKLAELKRKAQKIAIAIKKQLLSQMEK---EE 1664

Query: 350 MMKLELDLHGLREETS-AHIERKDKDITILQCRLOELQLEFTETQKL--TLKKDKFLQEKD 407  
K + H E N T + ++ ++ ++ IL+ +L+ ++ +ET + + K E++  
Sbjct: 1665 QYKKGTESH--LSELNTKLQEREREVHILEEKLSVSSSQSETLIVPRSAKNVAAYTEQE 1722

Query: 408 EM-----LQEL-EKKLTQVQNSLLKKEKEL-----EKQCCMATELEMTVK-EAKQDKSKE 455  
E +Q+ E+K++ +Q +L +KEK L EK++ +++ EM + + + K +  
Sbjct: 1723 EADSQGCYQKTYEEKISVLQRNLTEKEKLLQRVQGEKEETVSSHFEMRCYQERLIKLEH 1782

Query: 456 AECKAL--QAEVQKLNKNSLEAAQERLAAQQAQCK--EAAALAGCHLEDOTQKRLQKGL 511

Sbjct: 1783 AE K Q+ + L+ LEE ++ L Q + + + A +LE+ +QK L  
 AEAKQHEDQSMIGHLQEELEEKNKYSLIVAHVEKEGGKNNIQAQNLNVFDDVQKTL 1842  
 Query: 512 LLDKQKADTIQELQRELMQLQKESMAEKEQTSNRKRVEELS--LELSEALRKLENSDKE 569  
 ++K T Q L++++ L +S + ++ +R +EEL+ E +AL++++ +K  
 Sbjct: 1843 --QEKELTCQILEQKIKEL--DSCLVRQKEV-HRVEMEELTSKYKELQALQOMDGRNKP 1896  
 Query: 570 KRQLQKTVAEQD---MKMNDMLDRIKHQHQREQGSIKCKLEEDLQEA TKLLEDKREQLKK- 625  
 L++ E+ + +L ++ QH + E + Q+ K + ++ L+  
 Sbjct: 1897 TELLEENTEESKSHLVQPKLLSNMEAQHNLEFKLAGAEREKQKLGKEIVRLQKDLRML 1956  
 Query: 626 SKEHEKLMEGEALROEFKKKDKTLKENSRLKEENENLRAELQCCSTQLESSLNKYNT 685  
 KEH++ ELE L++E+ + E K+++E E+L EL+ ST L+ + ++NT  
 Sbjct: 1957 RKEHQ---ELEILKKEYDQ-----EREKIKQEQEOL--ELKHNT-LKQLMREFNT 2003  
 Query: 686 S-QQVIQDLNKEIALQKESLMSLQAQLDKALQEKHYLTITITKEAYDALSRKSAACQDD 744  
 Q Q+L I ++A+L ++ Q+E + L I E D L R +A  
 Sbjct: 2004 QLAQKEQELMENTIKETINKAQEVEAEELLESQETNQLLKXIA-EKDDDLKR-TAKRYEE 2061  
 Query: 745 LTOALEKLNHVTSETSKLOOSLTOTQEKKAQ-LEEIIAYEERMK--KLNTELRLKRGFH 801  
 A E+ +T++ + LQ L + Q+K Q LE+E + + +L T+L +  
 Sbjct: 2062 ILDAREE--EMTAKVRDLQTOLEELQKQYQKLEQENPGNDNVTIMELQTLAQKTTLI 2119  
 Query: 802 QESELEVHAFDKKLEEMSCQVLQWQK 827  
 +S+L+ F +++ + ++ ++K  
 Sbjct: 2120 SDSKLKEQEFREQIHNLEDRKKYK 2145  
 Score = 316 (47.4 bits), Expect = 7.1e-24, P = 7.1e-24  
 Identities = 213/977 (21%), Positives = 454/977 (46%)  
 Query: 4 EAGERD-REVSSLSNKLSLQLD-IKNLHDVCKRQRTLQDNQLCMEEAMNSSHOKKQK 61  
 E R+ +V S+ K L+ Q + ++ +H++ + Q K + +L + + ++ +  
 Sbjct: 1034 EVHRRELNDVISIWEKLNQQAELQEIHEI-QLQKEQEVAELKQKILLFGCEKEEMNK 1092  
 Query: 62 ALAFESEVEFGSSKQCHLRQLQ-QLKKLL---VLOQE--LEFTEELQTSYSLRQY 114  
 + + + E G + L +LQ QLK+K + Q E L+ H E+L+ +  
 Sbjct: 1093 EITLKKE---GVKQOTTLNELQELQKQSAHVNSLAQDETCLKAHLEKLEVDLNSLKE 1149  
 Query: 115 QSILEKQTSDLVLLHHCKLKEDEV---ILYEEEMGNHNTGKHLHAQELALAGDKI 171  
 + L++Q +L +L K K E+ + +E +++ EK + + +L K+  
 Sbjct: 1150 NTFLLQELVELKMLAEEDKRVSELTSLKTTDEEFQSLKSSHEKSNKSLDKSLFKKL 1209  
 Query: 172 AS-LERSNLVYRDKYQSSLS--NIELLECQVMKLGELGGIMGQEPENKGDHKSVRITYS 228  
 + L L++ K +L EL+ L I +++ K +  
 Sbjct: 1210 SEELAIQLDICCKTEALLEAKTNELINISSKTNAILSRI--SHCQHRTTKVKEALLK 1267  
 Query: 229 PCMIQEHQ-----ETQKRLSEVMQKVSQ-ODLIQELRNKLACSNAVLREKALIKL 280  
 C + E + E Q L+ +Q+ + Q ++ +++ A +LV E+E L  
 Sbjct: 1268 TCTVSELAQLRQLTEQNTLNISFQOATHQLEKENQIKSMKADIESLVTEKA---L 1323  
 Query: 281 QADFASCTATHRYPPSSSEECEDIKKILKHLQEQKDSQCLHVEEYQNLVKDLRVELEAVS 340  
 Q + + + S E C I ++ K L E ++ L E E +K+ +VE+ ++S  
 Sbjct: 1324 QKEGGN---QQQAASEKESC--ITQLKKESENINAVTLMKEE---LKEKKVEISSLS 1373  
 Query: 341 EQKRNIMKDMKLELDLHGLREETSABIERKDKDITILQCRQLQEL--QLEFTETQKLT-L 397  
 +Q ++ + + L S+ ++ D++ L ++Q+L +++ +K++ L  
 Sbjct: 1374 KQLTDLNVQLQN-SISLSEKAAISSLRKQYDEKCELLDQVQDLSFKVDTLSKERISAL 1432  
 Query: 398 KK-DKFLQEKDEMLQLEKKLTQVQNSLLKKEKELEKQOCCMATELEMTV---KEAKQDKS 453  
 ++ D + + E ++ + + TQ QN++ + +LE + A E + + KE ++  
 Sbjct: 1433 EQVDDWSNKFSEWKKKAQSRFTQHNTVKELQIQLELKSKEAYEKDEQINLLKEELDQON 1492  
 Query: 454 KEAECKALQAEVQKLKNSLEEAQKQERLAAQAAQCKEEAALAGCHLE-DTQKRLQKGLL 512  
 K +C + E K K +E+ + L +Q A + E + +E ++ ++ K  
 Sbjct: 1493 KRFDCLKGEMEDDKSMKMEKESNLETELKSTARIMELEDHITQKTIEISLNEVLKNY- 1551  
 Query: 513 LDKQKADTIQELQRELMQLQKESMAEKEQTSNRKRVEELSLELSEALRKLENSDKEKRO 572  
 ++QK +EL ++LQ Q+ + ++ L ++ +LE KE  
 Sbjct: 1552 -NQKNDIEHKELVQKLQHFQELGEEKDNRVKEAEKILTLENQVYSMAELETKKKLEH 1610  
 Query: 573 LQKTVAEQDMKNDMLDRIKHQHQREQ-GSIKCKLEEDLQEA TKL---EDKREQLKSK 627  
 + +V ++ ++ + DR++ + + +K K E+ + K L E+K EQ KK  
 Sbjct: 1611 VNLVSKSEELKALDRLESESAKLAELKRKAQKIAIKKQLLSQMEKEEQYKGT 1670  
 Query: 628 EHEKLMEGEALROEFKKKDKTLKENSRLKEE-ENENL---RAELQCCSTQLESSLNK 682  
 E EL OE ++ L+E + +E +E L A+ T+ E + ++  
 Sbjct: 1671 ESHL---SELNKLQEREREVHILEELKSVESSESSETLIVPRSAKNVAAYTEQEEADSQ 1727  
 Query: 683 ---YNTSQVIQDLNKEIALQKESLMSLQAQLDKALQEKHYLTITITKEAYDALSRKSA 739  
 T ++ I L + + +KE L+ Q +K H+ +E L A  
 Sbjct: 1728 GCVQKTYEEKISVLQRNLT-EKEKLLQRVGQ-EKEETVSSHFMRCQYQERLIKLEHAEA 1785

672



Sbjct: 2008 ++ L KE Q V + + Q TN Q K K+A EK + R  
KEQELMTIKETINKAQ-EVEAEELLES----QEETN--QLLK--KIA-EKDDDLKRTAK 2057

Query: 913 NYIAKLSGKDLHLSVMVHLQOENKKLKEIEKKMKAEN 952  
Y L ++ + + LQ + ++L+K+ ++K + EN

Sbjct: 2058 RYEEILDAREEEMTAKVRDLQTQLEELQKKYQKLEQEN 2097

Score = 300 (45.0 bits), Expect = 3.7e-22, P = 3.7e-22  
Identities = 195/961 (20%), Positives = 435/961 (45%)

Query: 1 MKDEAGERDREVSSLSKLLSLQLDIKN--LHDVCKRQRKTLQDNQLCMEAMSSSHDKK 58  
+KD+ + +N K L +LD+K L + +L+ +EE ++ D+

Sbjct: 657 LKDKIIFQAHIEMNEKTLE-KLDVKQTELESLSSELSEVLKARHK-LEELSVLKDQT 714

Query: 59 QAQALAFESEVEFGSSKQCHLRQLQKLKLLV-LQOEFHTEELQTSYSSLRQYQSI 117  
+E E + K H +Q+ +K+ V +Q+ + +++ L++

Sbjct: 715 DKMK--QLEAKMDEQKNHHQQQVDSIIKEHEVSIQTEKALKDQINQLELLKERDKH 771

Query: 118 LEKQTSDLVLLHHCKLKEDEVILYEEEMG---NHNENTGEKLHLAQEQLALAGDKIASL 174  
L++ + L K E E+ ++ ++ T E+ +EQLA K+ L

Sbjct: 772 LKEHQAHVENLEADIKRSEGEQQASAKLDVFSYQSATHEQTKAYEQALQQLKLLDL 831

Query: 175 ERSNLNRYDKYQSSLSNIELLECQVKMLQGEGLGIMGO-EPENKGDSKVRITYTSPCMIQ 233  
E L + + + + + + +M Q E +N KV+ T

Sbjct: 832 ETERILLTKQVAEVEAQKDVCTELDAHKIQQDLMOLEKQSEMEQKVKSLTQ-VYES 890

Query: 234 EHQETQKRLSEVWQKVSQDDLIQELRN---KLACSNAVLEREKALIKLQADFASCTA 289  
+ ++ K + Q + ++++I ++R ++ + +E ++ L ++ +

Sbjct: 891 KLEDGNKEQEQTKQILVEKENMILQREGQKKEITILTKLSAKEDSIHLNEEYET--- 947

Query: 290 THRYPPSSSEECEDIKKILKHLQEQKDSQCLHVEEYQNLVKDLRVELEAVSEQKRNIMKD 349  
++ + ++ E +K+ K +QE + L E L K+L +S++++

Sbjct: 948 --KFK-NQEKMEKVKQKAKEMQETLKKLLDQEA--KLKKELENTALELSQEKQKFNK 1002

Query: 350 MMKL-ELDLHGLREETSA-HIERKDKDITILQCRLEQLQLEFTETQKITLKKDKFLQEKD 407  
M+++ + + G+ + S +K++ ++ + +EL + +K ++ +LQE

Sbjct: 1003 LLEMAQANSAGISDAVSRLETNQKEQIESLTVHRRLENDVISIMEKLNQAAELQEIH 1062

Query: 408 EM-LQELEKKLTQVQNSLLK---KEKELEKQOCMATE----LEMTVKEAKQD-KSKEAC 458  
E+ LQE E+++ ++L +++E+ K+ E +T+ E++ K+ K A

Sbjct: 1063 EIQLQKEQEVAELKQKILLFGCEKEEMNKEITWLKEEGVKQDTTLNELQELQKQSAHV 1122

Query: 459 KALQAEVQKLNLSLEAKQOERLAAQAAQCKEAAAGCHLEDTOQKQLKGLLDDKQKA 518  
+L + K LK LE+ + + + + +E+ +E+ +R+ + L K K

Sbjct: 1123 NSLAQDETCLKAHLEKLEVDLNKSLKENTFLOEQLVELKMLAEEDKRRVSE--LTSKLKT 1180

Query: 519 OTIQELQRELQMLQKSSMAEKQTSNRKRVEELSLESEALRKLKLENSDKRQLQKTVA 578  
T +E Q +K + E + +K EEL++L +K E + K + +

Sbjct: 1181 -TDEEFQSLKSSHEKSNKSLKEDKSLFQKLLSELAIQLOICCKTEALLEAKTN--ELIN 1237

Query: 579 EQDMKMNMDLDRIKH-QHREQGSIKCKLEEDLQEQATKLEDKREQLKKSHEHEKLMGEL 637  
K N +L R I H QHR K++E L T + + QL++ E + +

Sbjct: 1238 ISSSKTNAILSRISHCQHRRT-----KVKEALLIKTCTVSELEAQLRQLTEEQNTLNISF 1292

Query: 638 EALRQEFKKKD---KTLKENSRLKEEENENLR-----AELQCCSTQLESSL---- 680  
+ + ++K+ K++K + L E E L+ +E + C TQL+ L

Sbjct: 1293 QOATHQLEEKENQIKSMKADIESLVTEKEALQKEGGNQQAASEKESCITQLKKELSENI 1352

Query: 681 NKYNTSQQVIQDLNKEIALQKESLSLQAQLDKALQ-KEKHYLQTTITKEAYDALSRKSA 739  
N ++ +++ EI+ + L L QL ++ EK +++ K+ YD +

Sbjct: 1353 NAVTLMKEELKEKKEISSLKQLTDLNVQLQNSISLSEKEAAISSLRKQ-YDEEKCELL 1411

Query: 740 ACQDDLTQALEKLN-HVTSETKSLQSLTQTEKKAQLEEEIIAYEERMKKLNTLR-KL 797  
DL+ ++ L+ S + + + E K + + ++ +K+L +L K

Sbjct: 1412 DQVQDLSFKVDTLSEKISALEQVDDWSNKFSEWKKQASRFTQHONTVKELQIQLELKS 1471

Query: 798 RGFHQSELEVHAFDCKLEEMSCQVLQWQKHONDLMKLAKEEQLR-EFQEEAALKEN 856  
+ +++ E +++ ++L++ + + + +D + KE L E + + A + E

Sbjct: 1472 KEAYEKDE-QINLKEELDQONRRFOCLKGEMEDDKSKMEKKESNLETLSQARIME- 1529

Query: 857 LLEDDKPCCLPQWSVPKDTCLRYRGNDQIMTNLEQWAKQKVANEKLGNLREQVNYIA 916  
LED + + T + N+ ++ N Q OK K +L +++ +

Sbjct: 1530 -LEDH-----ITOKTIEIESLNE-VLKNYNQ---QKDIENK---ELVQKLQHFQ 1570

Query: 917 KLSGKDH---LHSVMVHLQOENKKLKEIEKKMKAENTRLCTKA 959  
+L EKD+ ++ L+ + +K E+E KK + E+ L K+

Sbjct: 1571 ELGEEKDNRVKEAEKILTLENQVYMKAELETKKKELEHVNLSVKS 1617

Score = 298 (44.7 bits), Expect = 6.1e-22, P = 6.1e-22  
Identities = 207/886 (23%), Positives = 412/886 (46%)

Query: 47 MEEAMNSSHDKQAALAFEESEVEFGSSKQCHLRQLQQLKKLLVLQOELEFHTTELQT 106  
 + E N + + Q EE E + S K + L + LQ+E +  
 Sbjct: 1281 LTEEQNTLWISFQOATHQLEEKENQIKSMKA----DIESLVTEKALQKEGNGQQQAASE 1336

Query: 107 SYSLRQYQSILEKQTSIDLVLHHHCKLKEDEVILYEEEMGNHNTGEKHLAQEQLAL 166  
 + Q + L + + + L+ K K+ E+ + + + N + L+ + + A  
 Sbjct: 1337 KESCITQLKKELSENINAVTLMKEELKEKKVEISSLSKQLTDLNVQLQNSISLSEKEAA- 1395

Query: 167 AGDKIASLERSLNLRYDRKYSSLSNIELLECQVKNLQELGGIMGQEPENKGDHSHKVR 226  
 I+SL + Y + + L + + L + V L E + Q + S + +  
 Sbjct: 1396 ----ISSLRKQ---YDEEKCELLDQVQDLSFKVDTLSKEKISALEQVDDWSNKFSEWK-K 1447

Query: 227 TSPCMIQEHQETQKRLS-----EVWQKVSQODDLIQEL--RNK-LACSNALVLE--- 272  
 + +HQ T K L E + + K Q + L +EL +NK C + +  
 Sbjct: 1448 KAQSRFTQHQNTVKELQIQLELKSKEAYEKDEQINLLKEELDQONKRFDCCLKGEMEDDKS 1507

Query: 273 -REKALIKLOADFASCTAT---HRYPPSSSEECEDIKKILKHLQEQKDSQCLHVEEYQN 327  
 EK L + + S TA + + E E + +L+K+ +QKD E+ +  
 Sbjct: 1508 KMEKKESNLETLSQATARIMELEDHITQKTIEIESLNEVLKNYNQXDI-----EHKE 1561

Query: 328 LVKDLRVELEAVSEQKRNIMKMMKLELDLHGLREETSABIERKDKDI--TILQCRQLQEL 385  
 LV+ L+ + + E+K N +K+ + L L A + E K K+ + L + + E  
 Sbjct: 1562 LVQKLQ-HFQELGEEDNRVKEAEKILTLENQVSMKAELETKKKELEHVNLSVSKSEE 1620

Query: 386 QLEFTEQKLTLLKDKFLQEKDEMLOLEKKLTQVQNSLLKKEKELEKQOCMATELEMTV 445  
 +L+ E + L+ + + E+ + + E+K+ + + LL + + E+Q TE + +  
 Sbjct: 1621 ELKALEDR--LESES-AAKLAELKRAEKQIAAIAIKQLLSQMEKEEQYKKGTESHLSE 1676

Query: 446 KEAKQDKSKEAECKALQAEVQKLKNSLEEAQOERLAAQQAACK-EAALAGCHLEDQ 504  
 K + +E E L+ + + + +S E R A AA + EEA GC + +  
 Sbjct: 1677 LNTKLQE-REREVHILEEKLKSVSSQSETLIVPRSAKNVAAYTEQEEADSQGCVCQKTYE 1735

Query: 505 RKLQKGLLLDKQKADTIQELQRELQMLQKESMAEKEQTSNRKRVESLSEALRKLE 564  
 K+ +L + + + LQR Q +KE + + + + R + +E + +L A K  
 Sbjct: 1736 EKIS--VLQRNLTEKEKLLQRVQ--EKEETVSSHFM--RCQYQERLILKEHAQKQ 1788

Query: 565 NSDKERKLOKTVAEQDMKMDMLDRIKKHQHQREQ--SIKCK--LE---EDLQ-----E 611  
 LQ+ + E+ + K + + + H +E G +I+ K LE +D+Q E  
 Sbjct: 1789 EDQSMIGHLQEELEENKKNKYLIV--AQHVEKEGGKNIAQKQNLNVFDDVQKTLQEKE 1846

Query: 612 AT-KLEDKREQLKKSKEHEKLMEG-ELEALRQEFKKDKTLKENS---KLEENENL 665  
 T + +LE K + +L +K + E+ E L + + + + R +L EEN  
 Sbjct: 1847 LTCQILEQKIKELDSCLVRQKEVHRVEMEELTSKYELQALQMDGRNKPTELLEENTEE 1906

Query: 666 RAEQQCCSTQLESSLN-KYNTSQVQIDLNKEIALQKESLMSLAQLOKALQKEKHYLOT 724  
 + + +L S+ + +N + + + +E + + + +LQ L + L+KE H +  
 Sbjct: 1907 KSKSHLVQPKLLSNMEAQHNDLEFKLAGAEREKQKLGKEIVRLQKDL-RMLRKE-HQQL 1964

Query: 725 TITKEAYDALSRKSAACODDLTQALEKLNHVTSETSKLQSSLTQTOEKKAAQLEEEIAYE 784  
 I K+ YD R+ Q+ + LE L H + + + + + TQ +K+ +LE I +  
 Sbjct: 1965 EILKKEYDQ-EREKIKQEQ--EDLE-LKHNSTLKQLMREFNTQLAQKEQLEMTI---K 2017

Query: 785 ERMKKLNTLRLKRGFHOSELEVHAFDKKLEEMSCQVLOWQKQHQNDLKMALAAKEEQLR 844  
 E + K +L HQE E + KK+ E + + K+ + + +L A+EE+ +  
 Sbjct: 2018 ETINKAQEVEAEELSHQE---ETNQLLKKIAEKDDDLKRTAKRYE---EILDAREEEMT 2071

Query: 845 EFQEEMAALKENLLEDDKEPCCLPQWSVP-KDTCRLYRGNDQIMTNLEQWAKQOKVAN 903  
 + + E L + + + L Q P D + + + T L Q K + + + K  
 Sbjct: 2072 AKVRDLQTLQELQKYYQQK--LEQEENPGNDNVTIM---ELQTQLAQ--KTTLISDSK 2123

Query: 904 LGNQ-LREQVNYIA-KLSGEKDLHLSVMV-HL 932  
 L Q REQ+ + + +L + + + + + V HL  
 Sbjct: 2124 LKEQEFREQIHNLDRLLKKEKNVYATTVGHL 2155

Score = 280 (42.0 bits), Expect = 5.2e-20, P = 5.2e-20  
 Identities = 209/938 (22%), Positives = 432/938 (46%)

Query: 3 DEAGERDREVS-SLNSKLLSLQDLIKN-LHDVC-KRQRTLQDNQLCMEAM-NSSHDKK 58  
 + + + +E+ +L KLL + +K L + + +K Q N +E A NS+  
 Sbjct: 957 EKVQKAKEMQETLKKKLLDQEAKKKELENTALELSQKEQFNKMLEMAQANSAGISD 1016

Query: 59 QAALAFEESEVEFGSSKQCHLRQLQQLKKLLVLQOELEFHTTELQTSYSLRQYQSIL 118  
 L + E + S + H R+L + + + +L EELQ + + + +  
 Sbjct: 1017 AVSRLETNQKE-QIESLVEVHRRLENDV---ISIWKKLNQQAELQ-EIHEIQLQEK-- 1069

Query: 119 EKQTSDLV--LLHHHCKLKE-DEVILYEEEMGNHNTGEKHLAQEQLALAGDKIASLE 175  
 E+ + +L +L C+ +E + + I + +E G + T +L +Q + + +A E  
 Sbjct: 1070 EQEVAELKQKILLFGCEKEEMKEITWLKEEGVKQDTTLNELQELKQKSAHVNSLAQDE 1129

Query: 176 RSLNLYRDKYQSSLSNIELLECQVKNLQELGGI--MGQEPENKGDHSHKVRITYTSPCM 233  
 L + +K + L N L E LQ +L + + +E + K + +T+ Q

Sbjct: 1130 TKLKAHLEKLEVDL-NKSLKENT--FLQEQVLKMLAEEDKRVSELTSLKLTDEEFQ 1186

Query: 234 E---HQETQKRLSEVWQVVSQDDLIQELRNKL--AC--SNALVLEREKALIKLQADFA 285  
H+++ K L + K + L +EL +L C + AL+ + LI + +

Sbjct: 1187 SLKSSHEKSNKSLD---KSLEPKKLEELAIQLDICKKTEALLEAKTNELINISSKT 1243

Query: 286 SCTATH-RYPPSSSEECEDIKKILKHLQEQKDSQCLHVEEYQNLVKDLRVELEAVSEQRR 344  
+ + + + + + + + I + + + Q + E QN + + + E + K

Sbjct: 1244 NAILSRISHCQHRTTKVKEALLIKTCTVSELEAQLRLTEEQNTLNISFQQATHQLEEEK 1303

Query: 345 NIMKMMKLELD-LHGLREETSABIERKDKDITILQCLRLQELQLEFTET-QKLTLLKDKF 402  
N + K M K + + + L + E + + + + + + + + L + E + E + + TL K + +

Sbjct: 1304 NQIKSM-KADIESLVTEKEALQKEGNNQQAASEKSCITQLKKELSENINAVTLMKEE- 1361

Query: 403 LQKDEMLQELEKILTOVQNSLLKKEKELEKQOCMATELEMTVKEAKQDKSKEAECKALQ 462  
L + EK + L K + LT + N L + L + + + + L + E K + + + + L

Sbjct: 1362 LKEKKEVEISSKQLTDL-NVQLQNSISLSEKEAAISSLRKQYDEEKCELLDQVQ--DLS 1418

Query: 463 AEVQKLNKSLLEAKQOERLAAQQAQCKEEAALACCHLEDTORLQKGLLLDKQKA---- 518  
+ V L + A + Q + + + + K + A + + T + + LQ L L + A +

Sbjct: 1419 FKVDTLSEKISALEQVDDWSNKFSEWKKKAQSRFTOHONTVKELQIQLELKSKEAYEKD 1478

Query: 519 DTIQELQRELMQKESMAEKQTSNRKRVEELSLELSEALRKLENSDKERQKQKTVA 578  
+ I L + EL K + E + + + + L + L + + + + L + +

Sbjct: 1479 EQINLKEELDQKNRFDCLKGEDEDDKSKMEKKESNLET---ELKSQTARIMELEDHIT 1535

Query: 579 EQDMKMDMLDRIKHQHREQGSIKCK-LEEDLQEATKLLEDKREQLKKSHEKLMEGEL 637  
+ + + + + + + + K + + Q I + K L + LQ + L E + K + + K + + E + E + +

Sbjct: 1536 QKTIEIESLNEVLKN-YNOQKDIEHKELVQKLQHFQELGEEKDNVRKEAEKILTLENQV 1594

Query: 638 EALRQEFKKDKTLKENSRLKEEENENLRAELQCCSTQLES-SLNKYNTSQVQIDLNKE 696  
+ + + E + K K L + + + + E L + A L + + LES S K + + + + +

Sbjct: 1595 YSMKAELETKKLEHVNLSVKSKEEELKA-LE---DRLESESAKL---AELKRKAQK 1647

Query: 697 IALQKESLMSLAQDLKALQKEKHLYQTITKEAYDALSRSAACODDLTQALEKLNHVT 756  
IA K + L + S Q + + + KE + Y + T + + L + K + + + + EKL V

Sbjct: 1648 IAAIKQLLS---QME---EKEEQYKGT--ESHLSLNTKLQEREREVHILEEKLKSV 1699

Query: 757 S---ET---KSLQSLTOTQEKKAQLEEEII-AYEERMKKLNTLRLGRFHQSELEV 808  
S ET + S + T + + + A + + YEE + L L E E +

Sbjct: 1700 SSQSETLIVPRSAKNVAAYTEQEADSQCCVQKTYEEKISVLQRNLT-----EKEKLL 1752

Query: 809 HAFDKKLEMSQVQLWQKHQNDLKMALAEQQLREFQEEAALKENLLEDKPECCPL 868  
+ + EE + + Q + Q L L L E + E Q + L + E L E + K + +

Sbjct: 1753 QRVGQKEETVSSHFMRCQYQERLIKLEHAEAKQHEDQSMIGHQEELEENKKYSLIV 1812

Query: 869 QMSVPKDCRLRYRGNDQIMTNLEQ-WAKQOKVANER-LGNQLREQ-VNYIAKLSGEKDH 925  
V K + + N Q NLE + OK EK L Q + EQ + + + + +

Sbjct: 1813 AQHVEKEGGK---NNIAQAKNLENVDDVQKTLQEKELTCQILEQKIKELDSCLVRQKEV 1869

Query: 926 HSV-MVHLQEQENKKLK 940  
H V M L + + KL +

Sbjct: 1870 HRVEMEELTSKYEKLO 1885

Score = 227 (34.1 bits), Expect = 2.5e-14, P = 2.5e-14  
Identities = 160/716 (22%), Positives = 318/716 (44%)

Query: 233 QEHEQTKRLSEVWQVVSQDDLIQE-LRNKLACSNALV-LEREKALIKL-QADFASCTA 289  
+ E + TQ + + + V + L + + + L S + + L R + + L + D S TA

Sbjct: 53 RESGDTQSFAQLQLRVPSVESLFRSPIKESLFRSSSKESLVRTSSRESLNRLDLSSTA 112

Query: 290 THRYPPSSSEECEDIKKILKHLQEQKDSQCLHVEEYQNLVKDLRVELEAVSEQRRNIMKD 349  
+ P E ED+ L + + + Q L + + R + + + + + + +

Sbjct: 113 SFDPSSMDSEADLVGNSDSLKEQLIQRLR--RMERSLSYRGKSELVTAYQMLQRE 170

Query: 350 MMKLELDLHGLREETSABIERKDKDIT-ILQCLRLQELQLEFTETQKLTLLKDKFLQEKDE 408  
KL + G + + + + DK + I + R + ELQ + + + L + D L + EKD +

Sbjct: 171 KKKLQ----GILSQS-----QDKSLRRIAELE-EELQMDQAKKHLQEEFDASLEEKDQ 219

Query: 409 MLQLEKKILTQVQNSLLKKEKELEKQOCMATELEMTVKEAKQDKSKEAECKALQAE---V 465  
+ L + + + + L + + + + + LE + + + + E + + + + V

Sbjct: 220 YISVLQTVSLLKQLRLNGPMNVDLKPLP-QLEPQAEVFTKEENPESDGEPPVEDGTSV 278

Query: 466 QKLKNSLEAKQOERLA--AQQAAQC-KEEAALACCHLEDTORLQKGLL-LDKQKADTI 521  
+ L + + K + QE L + + Q KE+ L E Q + L + L + K K +

Sbjct: 279 KTLETQQRVKQENLLKRCKETIQSHKEQTLTSEKEALQELDERLQELEKIKDLHM 338

Query: 522 QELQRELMQKESMAEKQTSNRKRVEELSLELSEALRKLENSDKERQKQKTVAEQD 581  
E + + L + + E + + + E + + E L E + R K + Q

Sbjct: 339 AEKTKLITQLADAKNLIEQLEQDKGMVIAETKQRMHETLEMKEEIIAQLRSRIKQMTTQG 398

Query: 582 MKMNDMLDRIKHQHREOGSIKCKLEEDLQEAT-KLEEDKREQLK---KSKEHEKL-MEGE 636  
++ + ++ + E+ + +EA KL + EQ+K K+ E E++ ++ E  
Sbjct: 399 EELREQKEKSERAAFELEKALSTAQKTEEARRLKAEMDEQIKTIEKTSEERISLQQE 458

Query: 637 LEALRQEFKK-KDKTLKENSRLKEEENENLRAELQCCSTQLESSLNKYNTSQVQIDLNK 695  
L ++QE K+ +E KL++ +E EL +L L T ++ Q+ K  
Sbjct: 459 LSRVKQEVVDVMKSSSEEQIAKLQKLHEK---ELARKEQELTKKLQ---TREREFQEQMK 512

Query: 696 EIALQKESLMSLQAQLDKALQKEKHLYLQTTITKEAYDALSRKSAACQDDLTQALEKLN-H 754  
+AL+K L+ +K Q+ + + K+A S DL Q E  
Sbjct: 513 -VALEKSQSEYLLKISQEQEQESLALAELELQKKAILTESENKLR---DLQQEAETYRTR 568

Query: 755 VTSETKSLQQLSTQTEKKAQLEEEIIAYEERMKKLNTLRLKRGFHQSESELEV--HAFD 812  
+ SL++SL QE K Q ++ + E K N E+ + H+ +ELE H D  
Sbjct: 569 ILELESSLEKSL---QENKNQSKDLAVHLEAEKNKHNEITVMVEKHK-TELESKHKQD 624

Query: 813 KKLEEMSCQVLQWQKHONDLMKLAKEEQLE-----FOEEMAALKENLLED-DK 862  
E QVL+ +Q+Q +++ L K EQ +E FQ + + E LE D  
Sbjct: 625 ALWTE-KLQVLK--QQYQTEMEKLEKKEQEKETLLKDKEIIFQAHIEEMNEKTLEKLDV 681

Query: 863 EPCCLPQWSVPKDTCLRYRGNDQIMTNLEQWAKQKVANEKLGQNLREOVNIAKLSGEK 922  
+ L S+ + + + + L Q ++L ++ EQ N+ +  
Sbjct: 682 KQTELE--SLSSELSEVLKARHKLEELSVLKQDTDKMKQELEAKNDEQKNHHQQQVDSI 739

Query: 923 DHLHSMVMHLQEQENKKLKEIEEKKM 948  
H V + Q+ K LK +1 + ++  
Sbjct: 740 IKEHEVSI--QRTEKALKDQINQLEL 763

Score = 183 (27.5 bits), Expect = 1.3e-09, P = 1.3e-09  
Identities = 132/584 (22%), Positives = 251/584 (42%)

Query: 409 MLQELEKKLTQVQNSLLKKEKELEKQCCMATELEMTVKEAK-QDKSKEAECKALQAEVQK 467  
M ++L++K+++ Q L + + +T M + + + E + Q  
Sbjct: 1 MFKKLKQKISSEEQQLQALAPAQASSNSSTPTRMRSRTSSFTQLDEGTPNRESGDTQS 60

Query: 468 LKNSLE-EAKQQRERLAAQAAQCKEEAALAGCHLEDTQRKLQKGLLLDKQKA--DTIQEL 524  
L+ E L + ++ + + R+ L LD A D ++  
Sbjct: 61 FAQKLQLRVPSVESLFRSPIKESLFRSSSKESLVRTSSRESLNRDLDSSTASFDPPSDM 120

Query: 525 QRELQMLQKESMAEKEQTSNRKRVEELSL-----ELSEALRKLNSDKKRLQKTVAE 579  
E + L S + KEQ R R E SL + SE + + +EK++LQ +++  
Sbjct: 121 DSEAEDLVGNSOSLNKEQLIQLRLKERSLSSYRGKYSYELVTAYQMLQREKKKLQGLSQ 180

Query: 580 -QDMKNDMLDRIKHQHREOGSIKCKLEE---DLQEATK---LLEDKREQLKKSKEHEKL 632  
QD + + + + +Q + K EE L+E + +L+ + LK+ + +  
Sbjct: 181 SQDKSLRRIAELEELQMDQAKKHLQEEFDASLEEKDQYISVLQTQVSLKQLRNGPM 240

Query: 633 MEGEALRO-EFKKDKTLKENSRLKEE---ENENLRAELQCCSTQLESSLNKYNTSQ 688  
L+ L Q E + + T +EN E E+ L+ +++ N ++  
Sbjct: 241 NVDVLKPLPQLEPQAEVFTKEENPESDGPVVEDGTSVKTLETLOQRVQRNELLKRCKE 300

Query: 689 VIQDLNKEIALQKESLMSLQAQLDKALQKEKHLYLQTTITKEAYDALSRKSAACQDDLTQA 748  
IQ ++ L +LQ QLD+ LQ E ++ E +++ A +L +  
Sbjct: 301 TIQSHKEQCTLLTSEKALQEQLDERLQ-ELEKIKDLHMAEKTCLITQLRDA--KNLIEQ 357

Query: 749 LEK-LNHVTSETKSLQQLSTQTEKKAQLEEEIIAYEERMKKLNTLRLKRGFHQSELE 807  
LE+ V +ETK + + +T E K EEEI R+K++ T+ +LR Q+ + E  
Sbjct: 358 LEQDKGMVIAETK---RQMHELEMK---EEEIAQLRSRIKQMTTQGEELR--EQKEKSE 409

Query: 808 VHAFDKKLEEMSCQVLQWQKHONDLMKLAKEEQLEFQ-----EEMAALKENLLEDDE 863  
AF EE+ + QK + K+ A +EQ++ + EE +L++ L +E  
Sbjct: 410 RAAF---EELKALSTAQKTEEARRLKAEMDEQIKTIEKTSEERISLQQELSRVKQE 465

Query: 864 PCCLPQWSVPKDTCLRYRGNDQIMTNLEQ-WAKQKVANEKLGQNLQ-----EQVNYIAK 917  
+ S + + +L + +++ + EQ K+ + + Q++ Q Y+ K  
Sbjct: 466 VVDVMKSSSEEQIAKLQKLHEKELARKEQELTKLQTREREFQEQMKVALEKSQSEYL-K 524

Query: 918 LSGEKDHLHSMVMH-LQENKKLKEIEEK---KMKAEINTRLCTKALGPSRTESTOREK 972  
+S EK+ S+ + L+ + K+ E E K + +AE R L S +S Q K  
Sbjct: 525 ISQEQEQESLALAELELQKKAILTESENKLRDLQQAETYRTRILELESSLEKSLQENK 584

Pedant information for DKFZphtes3\_lg13, frame 1  
-----

Report for DKFZphtes3\_lg13.1

[LENGTH] 1007  
[MW] 117480.77  
[PI] 5.90

[HOMOL] TREMBL:AF092090\_1 product: "cp151"; Rattus norvegicus cp151 mRNA, partial cds.  
 O.0  
 [FUNCAT] 30.03 organization of cytoplasm [S. cerevisiae, YDL058w] 5e-15  
 [FUNCAT] 08.07 vesicular transport (golgi network, etc.) [S. cerevisiae, YDL058w] 5e-15  
 [FUNCAT] 09.10 nuclear biogenesis [S. cerevisiae, YDR356w] 1e-11  
 [FUNCAT] 30.04 organization of cytoskeleton [S. cerevisiae, YDR356w] 1e-11  
 [FUNCAT] 03.22 cell cycle control and mitosis [S. cerevisiae, YDR356w] 1e-11  
 [FUNCAT] 30.10 nuclear organization [S. cerevisiae, YKR095w] 1e-08  
 [FUNCAT] 11.04 dna repair (direct repair, base excision repair and nucleotide excision repair) [S. cerevisiae, YKR095w] 1e-08  
 [FUNCAT] 99 unclassified proteins [S. cerevisiae, YLR309c] 1e-08  
 [FUNCAT] 1 genome replication, transcription, recombination and repair [M. jannaschii, MJ1322] 4e-06  
 [FUNCAT] 09.13 biogenesis of chromosome structure [S. cerevisiae, YLR086w] 9e-06  
 [FUNCAT] 03.04 budding, cell polarity and filament formation [S. cerevisiae, YHR023w MYO1 - myosin-1 isoform] 3e-04  
 [FUNCAT] 08.22 cytoskeleton-dependent transport [S. cerevisiae, YHR023w MYO1 - myosin-1 isoform] 3e-04  
 [FUNCAT] 03.25 cytokinesis [S. cerevisiae, YHR023w MYO1 - myosin-1 isoform] 3e-04  
 [FUNCAT] 98 classification not yet clear-cut [S. cerevisiae, YJR134c] 5e-04  
 [EC] 3.6.1.32 Myosin ATPase 1e-16  
 [PIRKW] nucleus 3e-10  
 [PIRKW] phosphotransferase 6e-09  
 [PIRKW] duplication 2e-06  
 [PIRKW] citrulline 2e-12  
 [PIRKW] tandem repeat 1e-16  
 [PIRKW] endocytosis 2e-13  
 [PIRKW] heart 8e-13  
 [PIRKW] transmembrane protein 1e-13  
 [PIRKW] serine/threonine-specific protein kinase 6e-09  
 [PIRKW] zinc finger 2e-13  
 [PIRKW] metal binding 2e-13  
 [PIRKW] DNA binding 4e-12  
 [PIRKW] muscle contraction 1e-16  
 [PIRKW] acetylated amino end 1e-11  
 [PIRKW] actin binding 1e-16  
 [PIRKW] mitosis 5e-15  
 [PIRKW] microtubule binding 5e-15  
 [PIRKW] ATP 1e-16  
 [PIRKW] thick filament 1e-16  
 [PIRKW] phosphoprotein 4e-16  
 [PIRKW] skeletal muscle 2e-14  
 [PIRKW] calcium binding 2e-12  
 [PIRKW] alternative splicing 1e-16  
 [PIRKW] coiled coil 1e-16  
 [PIRKW] P-loop 1e-16  
 [PIRKW] heptad repeat 3e-10  
 [PIRKW] methylated amino acid 1e-16  
 [PIRKW] immunoglobulin receptor 2e-06  
 [PIRKW] peripheral membrane protein 2e-13  
 [PIRKW] cardiac muscle 8e-13  
 [PIRKW] hydrolase 1e-16  
 [PIRKW] microtubule 3e-10  
 [PIRKW] muscle 8e-13  
 [PIRKW] EF hand 2e-12  
 [PIRKW] cytoskeleton 2e-15  
 [PIRKW] hair 2e-12  
 [PIRKW] calmodulin binding 2e-13  
 [PIRKW] Golgi apparatus 3e-10  
 [SUPFAM] myosin heavy chain 1e-16  
 [SUPFAM] conserved hypothetical P115 protein 1e-07  
 [SUPFAM] centromere protein E 5e-15  
 [SUPFAM] unassigned Ser/Thr or Tyr-specific protein kinases 6e-09  
 [SUPFAM] calmodulin repeat homology 2e-12  
 [SUPFAM] myosin motor domain homology 1e-16  
 [SUPFAM] alpha-actinin actin-binding domain homology 2e-07  
 [SUPFAM] plectin 2e-07  
 [SUPFAM] trichohyalin 2e-12  
 [SUPFAM] pleckstrin repeat homology 8e-08  
 [SUPFAM] ribosomal protein S10 homology 2e-07  
 [SUPFAM] giantin 3e-13  
 [SUPFAM] protein kinase homology 6e-09  
 [SUPFAM] protein kinase C zinc-binding repeat homology 8e-08  
 [SUPFAM] kinesin motor domain homology 5e-15  
 [SUPFAM] human early endosome antigen 1 2e-13  
 [SUPFAM] M5 protein 1e-07  
 [PROSITE] LEUCINE\_ZIPPER 7  
 [PROSITE] MYRISTYL 2  
 [PROSITE] CAMP\_PHOSPHO\_SITE 2  
 [PROSITE] CK2\_PHOSPHO\_SITE 20

```

[PROSITE] TYR_PHOSPHO_SITE 1
[PROSITE] PKC_PHOSPHO_SITE 16
[PROSITE] ASN_GLYCOSYLATION 2
[KW] All_Alpha
[KW] LOW_COMPLEXITY 15.00 %
[KW] COILED_COIL 42.40 %

```

```

SEQ MKDEAGERDREVSSLSKLLSLQLDIKNLHDVCKRQRKTLQDNQLCMEEAMNSSHDKKQA
SEG .....XXXXXXXXXXXX.....
PRD ccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc
COILS .....CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC.....

SEQ QALAFESEVEFGSSKQCHLRQLQLKKLLVLQQLLEFTEELQTSYYSRLQYQSILEK
SEG .....XXXXXXXXXXXXXXX.....
PRD hhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh
COILS .....CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC.....

SEQ QTSDLVLLHHHCKLKEDEVILYEEEMGNHNTGEKHLAQEQLALAGDKIASLERSLNL
SEG .....
PRD hhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh
COILS .....CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC.....

SEQ YRDKYSSLSNIELLECQVKMLQGLGGIMGQEPENKGDHVKVRIYTSPCMIEHQETQK
SEG .....
PRD hhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh
COILS .....CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC.....

SEQ RLSEVWQKVSQQDDLIQELRNKLACSNAVLEREKALIKLQADFASCTATHRYPPSSSEE
SEG .....
PRD hhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh
COILS .....

SEQ CEDIKKILKHLQEQKDSQCLHVEEYQNLVKDLRVELEAVSEQKRNIMKDMMKLELDLHGL
SEG .....
PRD hhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh
COILS .....CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC.....

SEQ REETSAHIERKDKDITILQCLRLQELQLEFTETQKLTLLKDKFLQEKDEMLQLEKKLTQV
SEG .....
PRD hhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh
COILS .....CCCCCCCCCCCC.....

SEQ QNSLLKKEKELEKQQCMALETEMTVKEAKQDKSKEAECKALQAEVQKLNKSLSEAKQOER
SEG .....XXXXXXXXXXXX.....
PRD hhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh
COILS .....CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC.....

SEQ LAAQQAQCKEEAALAGCHLEDTQRKLQKGLLLDKQKADTIQELQRELQMLQKSSMAEK
SEG .....XXXXXXXXXXXX.....
PRD hhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh
COILS .....CCCCCCCC.....

SEQ EQTSNRKRVEELSLELSEALRKLNSDKKRLQKTVAEQDMKMDMLDRIKHQHREQGS
SEG .....XXXXXXXXXXXX.....
PRD hhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh
COILS .....CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC.....

SEQ IKCKLEEDLQEATKLEDKREQLKKSKEHEKMEGELEALRQEFKKDKTLKENSRLKEE
SEG .....XXXXXXXXXXXX.....
PRD hhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh
COILS .....CCCCCCCC.....

SEQ ENENLRAELQCCSTQLESSLNKYNTSQQVIQDLNKEIALQKESLMSLQAQLDKALQKEKH
SEG .....
PRD hhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh
COILS .....CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC.....

SEQ YLQTTITKEAYDALSRSAACQDDLTQALEKLNHVTSETKSLQOSLTQTEKKAQLEEEI
SEG .....XXXXXXXXXXXX.....
PRD hhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh
COILS .....CCCCCCCCCCCCCCCCCCCCCCCC.....

SEQ IAYEERMKKLWTELRLRGFQHESELEVHAFDKKLEEMSCQVLQWQKHQNDKMLAAKE
SEG .....
PRD hhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh
COILS .....CCCCCCCCCCCCCCCCCCCC.....

SEQ EGLREFQEEAALKENLLEDDKEPCCLPQWVSPKDTCLRYRGNDQIMTNLEQWAKQQKVA
SEG .....
PRD hhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh

```

Prosite for DKFZphtes3_ig13.1			
PS000001	52->56	ASN GLYCOSYLATION	PD0C00001
PS000001	684->688	ASN GLYCOSYLATION	PD0C00001
PS000004	240->244	CAMP_PHOSPHO_SITE	PD0C00004
PS000004	415->419	CAMP_PHOSPHO_SITE	PD0C00004
PS000005	74->77	PKC_PHOSPHO_SITE	PD0C00005
PS000005	110->113	PKC_PHOSPHO_SITE	PD0C00005
PS000005	238->241	PKC_PHOSPHO_SITE	PD0C00005
PS000005	290->293	PKC_PHOSPHO_SITE	PD0C00005
PS000005	392->395	PKC_PHOSPHO_SITE	PD0C00005
PS000005	399->399	PKC_PHOSPHO_SITE	PD0C00005
PS000005	444->447	EKC_PHOSPHO_SITE	PD0C00005
PS000005	503->506	PKC_PHOSPHO_SITE	PD0C00005
PS000005	544->547	PKC_PHOSPHO_SITE	PD0C00005
PS000005	566->569	PKC_PHOSPHO_SITE	PD0C00005
PS000005	600->603	PKC_PHOSPHO_SITE	PD0C00005
PS000005	650->653	PKC_PHOSPHO_SITE	PD0C00005
PS000005	655->658	PKC_PHOSPHO_SITE	PD0C00005
PS000005	735->738	PKC_PHOSPHO_SITE	PD0C00005
PS000005	876->879	PKC_PHOSPHO_SITE	PD0C00005
PS000005	968->971	PKC_PHOSPHO_SITE	PD0C00005
PS000006	39->43	CK2_PHOSPHO_SITE	PD0C00006
PS000006	53->57	CK2_PHOSPHO_SITE	PD0C00006
PS000006	68->72	CK2_PHOSPHO_SITE	PD0C00006
PS000006	116->120	CK2_PHOSPHO_SITE	PD0C00006
PS000006	190->194	CK2_PHOSPHO_SITE	PD0C00006
PS000006	250->254	CK2_PHOSPHO_SITE	PD0C00006
PS000006	296->300	CK2_PHOSPHO_SITE	PD0C00006
PS000006	439->443	CK2_PHOSPHO_SITE	PD0C00006
PS000006	444->448	CK2_PHOSPHO_SITE	PD0C00006
PS000006	471->475	CK2_PHOSPHO_SITE	PD0C00006
PS000006	520->524	CK2_PHOSPHO_SITE	PD0C00006
PS000006	536->540	CK2_PHOSPHO_SITE	PD0C00006
PS000006	566->570	CK2_PHOSPHO_SITE	PD0C00006
PS000006	576->580	CK2_PHOSPHO_SITE	PD0C00006
PS000006	650->654	CK2_PHOSPHO_SITE	PD0C00006
PS000006	674->678	CK2_PHOSPHO_SITE	PD0C00006
PS000006	804->808	CK2_PHOSPHO_SITE	PD0C00006
PS000006	888->892	CK2_PHOSPHO_SITE	PD0C00006
PS000006	963->967	CK2_PHOSPHO_SITE	PD0C00006
PS000006	968->972	CK2_PHOSPHO_SITE	PD0C00006
PS000007	135->143	TYR_PHOSPHO_SITE	PD0C00007
PS000008	207->213	MYRISTYL	PD0C00008
PS000008	599->605	MYRISTYL	PD0C00008
PS000029	93->105	LEUCINE ZIPPER	PD0C00029
PS000029	80->112	LEUCINE ZIPPER	PD0C00029
PS000029	90->119	LEUCINE ZIPPER	PD0C00029
PS000029	104->126	LEUCINE ZIPPER	PD0C00029
PS000029	403->425	LEUCINE ZIPPER	PD0C00029
PS000029	410->432	LEUCINE ZIPPER	PD0C00029
PS000029	918->940	LEUCINE ZIPPER	PD0C00029

group: cell structure and motility

DKF2phtes3.lkl1 encodes a novel 589 amino acid protein with strong similarity to *Mus musculus* actin-binding protein (ENC-1).

Ectoderm-neural cortex-1 protein (ENC-1) is an early and highly specific marker of neural induction in vertebrates. The protein is related to the kelch family proteins and is expressed during early gastrulation in the prospective neuroectodermal region of the epiblast and later in development throughout the nervous system (NS). ENC-1 functions as an actin-binding protein organising the actin cytoskeleton during neural differentiation and development of the NS. The novel protein is highly similar to ENC-1.

The new protein can find application in modulation of cyto skeleton organisation in human testicular cells.

strong similarity to mouse ENC-1

complete cDNA, complete cds, EST hits

Sequenced by DKF2

Locus: unknown

Insert length: 3525 bp

Poly A stretch at pos. 3515, polyadenylation signal at pos. 3499

```

1 GGTGGAGAGC CGGCCGACGG GAGCCGCGGC GGAGCCTGTT GAGCTCGCGC
51 GGGCTGCCGG GAGTGGTCTC TGAGGCGGCG GCGGCGGCGG GGATCGTCTC
101 CGGCACTGGC GCACCATGTC GGTCACTGTC CATGAGACCC GCAAGTCGCG
151 GAGCAGCAGC GGTCCATGTA ACGTCAACCTT CTTCACAAAG GCCTCCACCC
201 CGGACTGTGT GCTGGCCAC CTTCAACACG TTCGCAAGCA CTGCATGTTT
251 ACCGACGTCA CACTCTGGGC GGGCGACCGT GCCTTCCCTT GTACCCGTGC
301 CGTGCTGGCC GCCTCTAGCC GCTATTTTGA GGCCATGTTT AGCCATGGCC
351 TTCGGGAGAG CCGGGATGAC ACTGTCAACT TCCAGGACAA CCTGCACCCG
401 GAGGTGCTGG AGCTGCTGCT GGACTTTGCC TACTCTCAC GCATCGCCAT
451 CAAGCAGGAG AACGCTGAGT CACTGTGGA GCGAGGCGAC ATGCTCGCAT
501 TCCACCATGT GCGGGATGCT GCGGCGGAGT CCTTGAGAA GAACCTTTTC
551 CCTTCAACT GCCTGGCAT GATGCTCTC TCGGAGGCC ACCAGTGCCG
601 CCGGCTGTAT GAGTTCTCTT GCGCATGTG CTGGTGAC CTTGAGACGG
651 TGAGGCAGAG CGAGGACTTC AACAGCTGT CCAAGGACAC ACTGTGGAC
701 CTCATCTCGA GTGATGAGT GGAGACCGAG GACGAGCGGG TGGTCTTGA
751 GGCCATCCTC CAGTGGGTGA AGCAGCACT GGAGCCAGG AAGGTCCACT
801 TGCCCGAGCT CCTCCGAGC GTGCGTCTGG CCTTGCTGCC GTCCGACTGC
851 CTGCGAGGAG CCGTCTCCAG CGAGGCCCTC CTCATGGCAG ACGAGCGCAC
901 CAAGCTTATC ATGGATGAGG CCCTGCGCTG CAAGACGAG ATCTGCGAGA
951 ATGATGGCGT GGTCAACAGC CCCTGTGCCC GGCCACGCAA GCGGGGCCAC
1001 ACGCTACTCA TCCTGGGGGG CCAGACCTTC ATGTGTGACA AGATCTACCA
1051 GGTGGACAC AAGGCCAAGG AGATCATCCC CAAGCCGAC CTGCCAGGCC
1101 CCGGGAAGGA GTTCAGGCC TCAGCGATCG GCTGCAAGGT CTATGTGACG
1151 GGGGGCAGGG GCTCCGAGAA CGGGTCTCC AAGGATGCTT GGTGTGACGA
1201 CACCGTACAT GAGGAATGGT CCAAGCGCGC GCCATGCTGT ATTGCCCGCT
1251 TTGGCCATGG CTCAGCTGAG CTGGAGAACT GCCTCTATGT GGTGGGGGGA
1301 CACACATCCC TGGCAGGGGT CTTCGCCGCC TCGCCTCTGT TCTCCCTGAA
1351 ACAAGTGGAG AAATACGACC CTGGGGCCAA CAAGTGGATG ATGGTGGCCC
1401 CCTTGGCGGA TGGCGTCAGC AATGCCGAG TGGTGAATGC CAAGCTGAAG
1451 CTCTTTGTTT TCGGAGGAAC CAGCATCCAC CGGGACATGG TGTCCAAGGT
1501 CCAAGTGTAT GACCCCTCGG AGAACAGGTT GACGATCAAG GCGGAGTGCC
1551 CCCAGCCTTG GCGGTACACA GCGCTGCGG TCCTGGGCGC CCAGATCTTC
1601 ATCATGGGAG GTGACACGGA ATTCACAGCC GCTCGGCTT ACCGCTTTGA
1651 CTGTGAGACC AACCACTGGA CCGGATGAG GAGCATGACT GCCAAGGCGA
1701 TGCTCTGCA TCCTCTGCTT TCCGCAACA AGCTCTATGT GGTGCGGGGC
1751 TACTTTGGGA CCGAGAGGTG TAAGACTCTG GACTGTATG ACCCACTTC
1801 AGATACATGG AACTGCATCA CCACAGTGCC CTACTCACTT ATCCCCACGG
1851 CTTTGTGAG CACCTGGAAG CACCTGCCCG CGTGAGGAGC ACCTGCTGAG
1901 CCAAGCCAGA CCGCGCCCTT CAGTGTACA GCGTGGCCTT GCTTGTCTGC
1951 CACAGCGGGA GCTAAGCCGG CCCTGGGCCA GCACTCCGAG AGGTGGAAGG
2001 GGCCTTGCCA GCTCTGGGGA GCAGCAGCCT TGGGCTGTTT TGAGCTTAG
2051 GCAAGAGAAG AGAAGCATCT CTTGATCCG TGCCCTGGG GGCCTTTCA
2101 GCTTTGAGT GGTGTTGGG AAGACATACC TCCAGAGGG GCATGGACTG
2151 CCACCGAGAC TGACCTGGC GTGCGGAGA AGGACACTTG CAGAGCCTTG
2201 AGATCACCTG TTTGGCAGGT CCTGGACTGG GCGCGGCGAG GCAGGGGCG
2251 GGAGGCGCCC GGGGTGGGCT TTGGGGCTGC GGCATGCTAT CACATCCTTT
2301 CCCTCCTGGC CTGCCCTGCT GGGGCTCTAC TGCCATCTAT AGATGGTGT
2351 CTGGGCTGG GAAACTAGGT TCCAGGGGTT TGAGACAGA AAGGTGACCA
2401 AGACAGATTT TTTAAGTGC AGAACTGCA GGGGGGCTC AGTGACATCC
2451 ATGAGGCTTT ATTAGCAAG GACACCCAGA CCTCAAGGT TTGTGGGCCC
2501 CTCCACAAA GCTGTAAGTC CCAGCCACCC TACTCAGGGC CTGTCTCAGT
2551 GCTGTGGCCC GGTGGGGACA CAGTTGCTCG TGGCCACTCA GTGGAGCTGG
2601 GCCTGCAGCA GACTCAAGGC TCCGAGTGCC CTGGGGGTCA CCCCCTCCCT
2651 CCCCTCTCA GAGCCACCC TGAGAGGAGC CAGTGACCCC CATGGCACAC
2701 ACCTGCCAAC AGCACTGGGG GTTTCTCCC AGGAGACCAC GGTGCCCTCC

```



```

2751 AAGACCAGGA GCAGCTGTGA GCTGGAGACA GCAGAGGGAC CCCAGGGTGT
2801 CCCCTGCAGA TCCCACCAGG GCCGCATCCA TCTCAGTTGT GAGGACAGTG
2851 ACGGGACCTT CACCATCCCT TTGGGTTTTT GCCCCATTG GTCCTCTGAG
2901 CTCCAAGATA AGAATGGCCC CGAGAGAACT GCTGAACATT TGTTCAATTGC
2951 TGTCACTCTC TGAGTCACTG GGGTCCCTCA CCAGCACCTC CCTGACACCT
3001 GGGCTATGGA GAGGTGGCGC CCTGTCACTG ACCATCCTAA TGCTCTCGC
3051 TCACTCCCAA GCCACCATTT GAGAGGGAGG GGTGTGGTG CCCTGACAGG
3101 GACTGGGCAG GGTGTCCAAA CTTGGGGCTT CCCAGGCACC TGCAGTGTGA
3151 ACACTGCTTG GCTGGCTCAA GATTAGGGCC GCGAGGGGGG CTGTGCACAT
3201 ACCAGTTACT TAAGCAGCCA CGAGTGTCCC CCATGCCCTT GTGCGGGTCC
3251 TGGAGGCCTC TTGGGGGTGG GACCTTTGGG CAGGGTTTGC CCACTGACGC
3301 GCCCGCCATG GGGCACTGGC TGCATGGGGC TCCTTGGACC CTGTAGACCC
3351 AGCAGGAGCC TGGCCGCGGG GACTGCAGGG AGGGTGCCTG GACCCGTGGG
3401 GTTGCTTCAT TGAGATAAAG CACACTTATC ACATAGCACA AAGGACGTGC
3451 CATGGTGCTT TCCCCAAAAG TTGTGTTGCT TTTATCAGTT TTCTAACTTA
3501 ATAAAAAGAG TTGAGAAAAA AAAAA

```

## BLAST Results

No BLAST result

## Medline entries

98350113:  
Cloning of human ENC-1 and evaluation of its expression  
and regulation in nervous system tumors.

97252647:  
ENC-1: a novel mammalian kelch-related gene specifically expressed in  
the nervous system  
encodes an actin-binding protein.

98234394:  
NRP/B, a novel nuclear matrix protein, associates with  
p110(RB) and is involved in neuronal differentiation

## Peptide information for frame 2

ORF from 116 bp to 1882 bp; peptide length: 589  
Category: strong similarity to known protein  
Classification: Cell structure/motility

```

1 MSVSVHETRK SRSSTGSMNV TLFHKASHPD CVLAHLNLT LR KHCMTDVT L
51 WAGDRAFPCH RAVLAASSRY FEAMFSGHLR ESRDDTVNFQ DNLHPEVLEL
101 LLDFAYSRI AINEENAESL LEAGDMLQFH DVRDAAAEFL EKNLFPSNCL
151 GMMLSDAHQ CRRLEYFSWR MCLVHFETVR QSEDFNSLSK OTLLDLISSD
201 ELETEDERVV FEAILQWVKH DLEPRKVHLP ELLRSVRLAL LPSDCLQEAV
251 SSEALLMADE RTKLIMDEAL RCKTRILQND GVVTSPCARP RKAGHTLLIL
301 GGQTFMCDKI YQVDHKAKEI IPKADLPSPR KEFSASAIGC KVVYTGGRGS
351 ENGVS KDVMV YDTVHEWSK AAPMLIAREG HGSAELENCL YVVGHTSLA
401 GVFPASPSVS LKQVKEYDPG ANKWMHVAFL RDGVSNAVV SAKLKLFPVG
451 GTSIHRDMVS KVQCYDPSEN RWTIKAECPQ PWRYTAAAVL GSOIFIMGCD
501 TEFTAASAYR FDCETNQWTR IGDHTAKRMS CHALASQNL YVVGGYFGTQ
551 RCKTLDICYD TSQTNWNCIT VPYSLIPTAF VSTWKHLPA

```

## BLASTP hits

Entry MMU65079\_1 from database TREMBL:  
gene: "ENC-1"; product: "actin-binding protein"; Mus musculus  
actin-binding protein (ENC-1) mRNA, complete cds.  
Score = 2402, P = 1.9e-249, identities = 440/589, positives = 513/589

Entry AF059611\_1 from database TREMBLNEW:  
gene: "NRPB"; product: "nuclear matrix protein NRP/B"; Homo sapiens  
nuclear matrix protein NRP/B (NRPB) mRNA, complete cds.  
Score = 2400, P = 3.0e-249, identities = 440/589, positives = 512/589

Entry AF010314\_1 from database TREMBL:  
gene: "PIG10"; product: "Pig10"; Homo sapiens Pig10 (PIG10) mRNA,  
complete cds.  
Score = 1745, P = 7.8e-180, identities = 335/507, positives = 403/507

Entry KELC\_DROME from database SWISSPROT:  
RING CANAL PROTEIN (KELCH PROTEIN). >TREMBL:DMRCPA\_1 product: "ring canal protein"; Drosophila melanogaster ring canal protein and ORF2 mRNA, complete cds.  
Score = 672, P = 3.9e-66, identities = 168/536, positives = 257/536

Alert BLASTP hits for DKFZphtes3\_1kl1, frame 2

No Alert BLASTP hits found

Pedant information for DKFZphtes3\_1kl1, frame 2

Report for DKFZphtes3\_1kl1.2

(LENGTH) 589  
(MW) 65923.45  
(PI) 6.10  
(HOMOL) TREMBL:MMU65079\_1 gene: "ENC-1"; product: "actin-binding protein"; Mus musculus  
actin-binding protein (ENC-1) mRNA, complete cds. 0.0  
(FUNCAT) 10.05.99 other pheromone response activities [S. cerevisiae, YHR158c]  
2e-09  
(BLOCKS) BL01016D Glycoprotease family proteins  
(PIRKW) zinc finger 1e-08  
(PIRKW) DNA binding 1e-08  
(PIRKW) transcription factor 1e-08  
(SUPFAM) POZ domain homology 3e-68  
(SUPFAM) vaccinia virus 59K HindIII-C protein 1e-15  
(SUPFAM) A55R protein 5e-29  
(SUPFAM) hypothetical protein YHR158c 4e-08  
(SUPFAM) A55R protein middle region homology 5e-29  
(SUPFAM) myxoma virus M9-R protein 1e-14  
(SUPFAM) A55R protein carboxyl-terminal homology 5e-29  
(KW) Alpha\_Beta

SEQ MSVSVHETKRSSTGSMNVTLFHKASHPDCVLAHLNLTLRKHCMTDVTWAGDRAFPCH  
PRD ccc

SEQ RAVLAASSRYFEAMFSGHGLRESRDDTVNFQDNLHPEVLELLDFAYSSRIATINEENAESL  
PRD hcc

SEQ LEAGDMLQFHQVDRDAAAEFLKKNLFPSNCLGMMLLSDAHQCRRLYEFSSWRMCLVHFETVR  
PRD hhh

SEQ QSEDFNSLSKDTLLDLISSDELETEDERVVFETAILQWVKHOLEPRKVHLPPELLRSVRLAL  
PRD hhh

SEQ LPSCQLQEAVSSEALLMADERTKLIMDEALRCKTRILQNDGVVTPCARPRKAGHTLLIL  
PRD cchhh

SEQ GGQTFMCDKIYQVDHKAKEIIPKADLPSPRKEFSASAIGCKVYVTGGRGSENGVSKDVWV  
PRD ccc

SEQ YDTVHEEWSKAAPMLIARFGHGSAELENCLYVVGHTSLAGVFPASPSVSLKQVEKYDPG  
PRD ccc

SEQ ANKMMMVAPLRDGVSNAAVVSAKLKFVFGGTSIHRDMVSKVQCYDPSNRWTIKAECPQ  
PRD ccc

SEQ PWRYTAAAVLGSQIFIMGDTPTAASAYRFDCTNQWTRIGDMTAKRMSCHALASGNKL  
PRD ccc

SEQ YVVGGVFGTQRCKTLDCYDPTSDTWNCTITVPYSLIPTAFVSTWKHLPA  
PRD ecc

(No Prosite data available for DKFZphtes3\_1kl1.2)

(No Pfam data available for DKFZphtes3\_1kl1.2)

DKFZphtes3\_in3

group: signal transduction

DKFZphtes3\_in3 encodes a novel 1196 amino acid protein with similarity to S. pombe Tup1 protein.

The protein contains 1 WD-40 repeat, which is typical for the beta-transducin subunit of G-proteins. The beta subunits seem to be required for the replacement of GDP by GTP as well as for membrane anchoring and receptor recognition. In addition, a RGD site is present.

The new protein can find application in modulating/blocking G-protein-dependent pathways.

similarity to Tup1p

complete cDNA, complete cds, EST hits

Sequenced by DKFZ

Locus: /map="6q24"

Insert length: 5277 bp

Poly A stretch at pos. 5267, polyadenylation signal at pos. 5244

```
1 GCTGCATAAA GCTGAGAGAT GCCTACAGCT GAGAGTGAAG CAAAAGTAAA
51 AACCAAAGTT CGCTTTGAAA AATTGCTTAA GACCCACAGT GATCTAATGC
101 GTGAAAAGAA AAAACTGAAG AAAAAACTTG TCAGGTCTGA AGAAAACATC
151 TCACCTGACA CTATTAGAAG CAATCTTCAC TATATGAAAG AAACACAAAG
201 TGATGATCCC GACACTATTA GAAGCAATCT TCCCCTATT AAAGAACTA
251 CAAGTGATGA TGAAGTGTCT GCTAACACTA ACAACCTGAA GAAGAGCACG
301 AGAGTCACTA AAAACAAATT GAGGAACACA CAGTTAGCAA CTGAAAATCC
351 TAATGGTGAT GCTAGTGTAG AGGAAGACAA ACAAGGAAAG CCAAAATAAA
401 AGGTGATAAA GACGGTGCCC CAGTTGACTA CACAAGACCT GAAACCGGAA
451 ACTCCTGAGA ATAAGGTTGA TTCTACACAC CAGAAAACAC ATACAAAGCC
501 ACAGCCAGGC GTTGATCATC AGAAAAGTGA GAAGGCAAT GAGGGAAGAG
551 AAGAGACTGA TTTAGAAGAG GATGAAGAT TGATGCAAGC ATATCAGTGC
601 CATGTAACCTG AAGAAATGGC AAAGGAGATT AAGAGGAAA TAAGAAAGAA
651 ACTGAAGAGAA CAGTTGACTT ACTTTCCCTC AGTACTTTA TTCCATCATG
701 ACAAACTAAG CAGTGAAAAA AGGAAAGAG AAGAGGAGT TCCAGTCTTC
751 TCTAAAGCTG AACAAGTAC ATTGACCATC TCTGGTGACA CAGTTGAAGG
801 TGAACAAAG AAAGAATCTT CAGTTAGATC AGTTTCTTCA GATTCTCATC
851 AAGATGATGA AATAAGCTCA ATGGAACAAA GCACAGAAGA CAGCATGCAA
901 GATGATACAA AACCTAAACC AAAAAAACA AAAAAGAAGA CTAAAGCAGT
951 TGCAGATAAT AATGAAGATG TTGATGGTGA TGGTGTTCAT GAAATAACAA
1001 GCCGAGATAG CCCGGTTTAT CCCAAATGTT TGCTTGATGA TGACCTTGTC
1051 TTGGGAGTTT ACATTACCGC AACTGATAGA CTTAAGTCAG ATTTTATGAT
1101 TTCTCACCCA ATGGTAAAAA TTCATGTGGT TGATGAGCAT ACTGGTCAAT
1151 ATGTCAAGAA AGATGATAGT GGACGGCCTG TTTTCATCTT CTATGAAAAA
1201 GAGAATGTGG ATTATATTCT TCCTATTATG ACCCAGCCAT ATGATTTTAA
1251 ACAGTTAAAA TCAAGACTTC CAGAGTGGGA AGAACAAATT GTATTTAATG
1301 AAAATTTTCC CTATTGTGCT CGAGGCTCTG ATGAGAGTCC TAAAGTCATC
1351 CTGTCTTTTG AGATTCTTGA TTTCTTAAGC GTGGATGAAA TTAAGAATAA
1401 TTCTGAGGTT CAAAACCAAG AATGTGGCTT TCGGAAAAAT GCCTGGGCAT
1451 TTCTTAAGCT TCTGGGAGCC AATGGAAATG CAACATCAAA CTCAAAACTT
1501 CGCTTGAGCT TATATTACCC ACCTACTAAG CCTCGATCCC CATTAAAGTG
1551 TGTGAGGCA TTTGAATGCT GGTCAAAATG TCCAAGAAAT CATTACCCAT
1601 CAACACTGTA CGTAACCTGA AGAGGACTGA AAGTTCCAGA CTGTATAAAG
1651 CCATCTTACC GCTCTATGAT GGCTCTTCAG GAGGAAAAAG GTAAACCACT
1701 GCATTGTGAA CGTCACCATG AGTCAAGCTC AGTAGACACA GAACCTGGAT
1751 TAGAAGAGTC AAAGGAAGTA ATAAAGTGGG AACGACTCCC TGGCGAGCT
1801 TGCCGATATCC CAAACAAACA CCTCTTCTCA CTAATGTCAG GAGAACGAGG
1851 ATGTTTTTGT CTGATTCTCT CCCACATGGG AAGAATATTA GCAGCAGCTT
1901 GTGCCAGCCG GGATGGATAT CCAATTATTT TATATCAAAAT TCCTTCTGGA
1951 CTTTTCATGA GAGATTTGTG TGGCCACCTC AATATCATTT ATGATCTTTC
2001 CTGCTCAAAA GATGATCACT ACATCCTTAC TTCATCATCT GATGGCACTG
2051 CCAGGATATG GAAAAATGAA ATAAACAATA CAAATACCTT CAGAGTTTTA
2101 CCTCATCCTT CTTTGTGTTA CACGGCTAAA TTCCATCCAG CTGTAAGAGA
2151 GCTAGTAGTT ACAGGATGCT ATGATTCCAT GATACGGATA TGGAAAGTTG
2201 AGATGAGAGA AGATTCTGCC ATATTGGTCC GACAGTTTGA TGTTCACAAA
2251 AGTTTTATCA ACTCACTTTG TTTTGATACT GAAGGTCATC ATATGTATTC
2301 AGGAGATTGT ACAGGGGTGA TTGTTGTTTG GAATACCTAT GTCAAGATTA
2351 ATGATTGGGA ACATTCAGTG CACCACTGGA CTATAAATAA GGAAATTAAG
2401 GAAACTGAGT TTAAGGGAAT TCCAATAAGT TATTGGAGA TTCAATCCAA
2451 TGGAAACGCT TTGTTAATCC ATACCAAGA CAGTACTTTG AGAATTATGG
2501 ATCTCCGGAT ATTAGTAGCA AGGAAGTTTG TAGGAGCAGC AAATTATCGG
2551 GAGAAGATTG ATAGTACTTT GACTCCATGT GGGACTTTTC TGTTCGCTGG
2601 AAGTGAGGAT GGTATAGTGT ATGTTTGGAA CCCAGAAACA GGAGAACAAG
```

2651 TAGCCATGTA TTCTGACTTG CCATTCAAGT CACCCATTGG AGACATTCTT  
2701 TATCATCCAT TTGAAAATAT GGTTCGATTC TGTCATTTC GGCAAAATGA  
2751 GCCAATTCAT CTGTATATTT AGGATTTCCA TGTTGCCAG CAGGAGGCTG  
2801 AAATGTTCAA ACCTTACAAT GGAACATTTT CATTACCTGG AATACACCAA  
2851 AGTCAGATG CCCTATGTAT CTGTCCAAA CTACCCCATC AAGCTCTTT  
2901 TCAGATTGAT GAATTTGTCC ACACCTGAAG TTCTTCAACG AAGATGCAGC  
2951 TAGTAAACA GAGGCTTGA ACCTGCACAG AGGTGATACG TTCTGTGCT  
3001 GCAAAAGTCA ACAAATCT CTCTTTACT TCACCACAG CAGTTTCCTC  
3051 ACAACAGTCT AAGTTAAAGT AGTCAACAT GCTGACCGCT CAAGAGATTC  
3101 TACATCAGTT TGGTTTCACT CAGACCGGGA TTATCAGCAT AGAAGAAAG  
3151 CCTTGTAAAC ATCAGGTAGA TAGCAGCACA ACGGTAGTGG CTCTTTATGA  
3201 CTACACAGCG AATCGATCAG ATGAACTAAC CATCCATCGG GGAGACATTA  
3251 TCCGATGTT TTTCAAAGAT AATGAAGACT GGTGGTATGG CAGCATAGGA  
3301 AAGGGACAGG AAGGTTATTT TCCAGCTAAT CATGTGGCTA GTGAAACACT  
3351 GTATCAAGAA CTGCCTCCTG AGATAAAGGA GCGATCCCTC CCTTTAAGCC  
3401 CTGAGGAAAA AACTAAAATA GAAAAATCTC CAGCTCCTCA AAAGCAATCA  
3451 ATCAATAGA ACAAGTCCCA GGACTTCAGA CTAGGCTCAG AATCTATGAC  
3501 ACATTCTGAA ATGAGAAAAG AACAGAGCCA TGAGGACCAA GGACACATAA  
3551 TGGATACAGG GATGAGGAAG AACAGCAAG CAGGCAGAAA AGTCACCTTA  
3601 ATAGAGTAAA GAATTGAAGA AAAGTTAAGA GCTGCCGAAA TGCACAGAGG  
3651 TGAATAAGAC AAACCAATG GAATTTCTCT TCAGAGTTCA GAATTTTCTCAG  
3701 ATACTAAGGA GGAAGAAAGG ATCCACTACT TCTTGTCTTT ATGAATGACT  
3751 CTAGAAAAAT CAGAATCAAG TTGTGGGGGG AAAATCAAC GTGGCCTTGG  
3801 AGTTCAGTTG TTATAACCA TTGTGACTAT TGTTGGTCAA AGTATTGGTA  
3851 CTATATTTCT TAGTAATGCG ATCATAATTA CATTACCACT GTTGGAAAAAC  
3901 TAATGAAGAA AACACTGTAA TTGCTACTCA GCAATGTGA ATAAAGGTG  
3951 TTTGCGTTAT TAGGATGCT GTTAAGTAAT CATTAAATAT TATTATATTG  
4001 GTAATGGTTG TATGTGTGAT GCTATGCCCA GAATATGAAG TATCTGTTTT  
4051 TGAATTCAC TTTATTTAAA AGATAAGCAG CTGACTGGGC ACGGTGCCTC  
4101 ATGCCTGTAA TCCTGACACC TTGGGAGGCT GAGGCAGGTG GATCACCTAA  
4151 GGTGAGGAGT TCAACAACAC CAGCCTGACC AACATGGTGA AACCCCATCT  
4201 CTACTAAAAA TACAATAATC AGCCGGGCTC CATGGCAGGC ACCTGTAATC  
4251 CCATCTACTG AGGCAGGAGA ATTGCTTGAC CCAGGAGGCA GAGGTTGCAG  
4301 TGAGCCAAAG TCACGCCATT GCACCTCCAG CTGGGGGACA GAGCAAGACT  
4351 CTATCTCCAA AAAACAAAAA AGATAAGCAG CTTAGAAATA TGGCGCATTC  
4401 AAAACAGTCT CAGTAACAAA GACATTAATA GAAAAAAT TACTTTCTAA  
4451 TTAATAATTT GTGTTCTTAA AGATCAAAAT ATATAGGTAA CTTTATAGAC  
4501 CTAAATTAAG AGTGATTTT GGTGGAAGT GCAACAATG TCCCAATGTC  
4551 TTTACTTTTT AAAAAGGCT TTTATATTT AAGCACATAC CTATTTTGTG  
4601 GACTTACATT GTTTAATATT TATTTTAATC TTAATATTTT TACATTATTA  
4651 TATTGCATTA TTTATTTTTT CTAAGTTCCA GAATAATAGT TCATTATTA  
4701 TAGACTATAT GTTTGAGCT TTGATATTAT AATGGGATAT TCATTTTTTG  
4751 TCTTTTCTT GACTCCTTTC TCAAGTGTGT GATAAGGTCT GCTGATAAAA  
4801 TATTTAAGCC CAAGAAAGTG AAACTAATA TAAATTTAGA AAGACCTATC  
4851 CAAATAGAC AGTCAATTC ATTAATAATA GAAGTGAGAA AAACAATGTT  
4901 GGCATTGAG GTGTAAATTT TGCCAGATG TATACCCAGT GTGAAATATC  
4951 TTCTAATAAA AATATATTTG GCTCTTATCC CTGCACATGT AGAGGCATAA  
5001 AAATTGGTAA ACATGTCCCG CTGTGTAGAA CTTTAAAAA AAGGCATTTT  
5051 TGAAGTGTGT GAGTGGCAGT GATAACTGGT GAAGCCTACA GCCATCCGCC  
5101 CAAAAGTCTG TTCTGATGCG ACTGAGTTTT CATTGTTCTG GATGTATAAG  
5151 TCTGTGTGTC AGGTACAGCT GGGCCAGGCC AGCTTGAGTC ACTCTTGATC  
5201 AAGCTTGTTT TTTTCTGCT TGTGAATGCA CTTGATAATT TAAAAATAAA  
5251 AATATCTGTT TCTCTGAAA AAAAAA

## BLAST Results

Entry HS32B1 from database EMBL:  
Human DNA sequence \*\*\* SEQUENCING IN PROGRESS \*\*\* from clone 32B1  
Score = 4445, P = 0.0e+00, identities = 889/889

Entry U93816 from database EMBL:  
Human exon-trapped sequence from 6q24.  
Score = 965, P = 4.0e-35, identities = 193/193

## Medline entries

No Medline entry

## Peptide information for frame 1

ORF from 19 bp to 3606 bp; peptide length: 1196  
Category: similarity to known protein

```

1 MPTAESEAKV KTKVRFKELL KTHSDLMREK KKLKKLVRS EENISPDITR
51 SNLHYMKETT SDDPDTIRSN LPHIKETTSO DVSAANTNNL KKSTRVTNKK
101 LRNTQLATEN PNGDASVEED KQKPNKKVI KTVPQLTTOD LKPETPENKV
151 DSTHQKTHTK PQPGVDHQS EKANEGRRET DLEDEELMQ AYQCHVTEEM
201 AKIIRKXIRK KLKEQLTYFP SOTLFHDDKL SSEKRRKKKE VPVFSKAETS
251 TLTISGDTVE GEQKKESSVR SVSSDSHDD EISSMEQSTE DSMQDDTKPK
301 PKTKKKKTKA VADNNEDVDG DGVEITSRD SPVYFKCLLD DDLVLGVYIH
351 RTDRLKSDFM ISHPMKIHW VDEHTGQYVK KDDSGRPVSS YYEKENVDYI
401 LPIMTQPYDF KQLKSRLPEW EEQIVFNENF PYLLRGSDES PKVILFFEIL
451 DFLSVDEIKN NSEVQNECG FRKIWAFLK LLGANGNANI NSKLRLQLYY
501 PPTKPRSPLS VVEAFWWSK CPRNHYPSTL YVTVRGLKVP DCIKPSYRSM
551 MALQEEKGKP VHCERHHESS SVOTEPGLEE SKEVIKWRRL PGACRIPNK
601 HFLSLNAGER GCFCLDFSHN GRILAAACAS RDGYPIILYE IPSGRFMREL
651 CGHLNIIYDL SWSKDDHYIL TSSSDGTARI WKNEINNTNT FRVLPHPSFV
701 YTAKEFPAVR ELVVTGCYDS MIRIWKVEMR EDSAILVRQF DVHKSFINSL
751 CFDTGHHMY SGCCTGVIVV WNTYVKINDL EHSVHHWTIN KEIKETEFKG
801 IPISYLEIHP NGKRLLIHTK DSTLRIMDLR ILVARKFVGA ANIREKIHST
851 LTPCGTFLFA GSEDGIVYVW NPETGEQVAM YSDLFFKSP1 RDISYHPFN
901 HVAFCAGQW EPILLIYDF HVAQGEAMF KRYNGTFLP GIHQSQDALC
951 TCRKLPHQGS FOIDEFVHTE SSSTKMLVK QLETVTEVI RSCAAKVNKN
1001 LSFTSPPAVS SQSKLKQSN MLTAQELHQ FGFTOTGIIS IERKPCNHQV
1051 DTAPTVALY DYTANRSEDL TIHRGDIRV FFKDNEDWYV GSIGKQGEY
1101 FPAHVASET LYQELPPEIK ERSPLSPPEE KTKIEKSPAP OKQSINKNKS
1151 QDFRLGSESM THSEMRKEQS HEDQGHIMDT RMKKNKQAGR KVTLIE

```

## BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphtes3\_ln3, frame 1

TREMBL:U92792\_1 gene: "tup1"; product: "Tup1"; Schizosaccharomyces pombe general transcriptional repressor Tup1 (tup1) mRNA, complete cds., N = 1, Score = 186, P = 1e-10

TREMBL:AF104258\_1 gene: "Pmc733"; product: "putative copper-inducible 35.6 kDa protein"; Festuca rubra putative copper-inducible 35.6 kDa protein (Pmc733) mRNA, complete cds., N = 1, Score = 235, P = 4.6e-18

TREMBL:SPAC3H5\_8 gene: "SPAC3H5.08c"; product: "beta-transducin"; S.pombe chromosome I cosmid c3H5., N = 2, Score = 231, P = 2e-14

PIR:T02533 hypothetical protein F13M22.17 - Arabidopsis thaliana, N = 2, Score = 228, P = 1e-13

TREMBL:AF104258\_1 gene: "Pmc733"; product: "putative copper-inducible 35.6 kDa protein"; Festuca rubra putative copper-inducible 35.6 kDa protein (Pmc733) mRNA, complete cds., N = 1, Score = 235, P = 4.6e-18

TREMBL:SPAC3H5\_8 gene: "SPAC3H5.08c"; product: "beta-transducin"; S.pombe chromosome I cosmid c3H5., N = 2, Score = 231, P = 2e-14

TREMBL:CERO3E1\_1 gene: "R03E1.1"; Caenorhabditis elegans cosmid R03E1, N = 1, Score = 215, P = 2.3e-13

SWISSPROT:YZLL\_CAEEL HYPOTHETICAL 43.1 KD TRP-ASP REPEATS CONTAINING PROTEIN K04G11.4 IN CHROMOSOME X., N = 1, Score = 203, P = 7.1e-13

>TREMBL:AF104258\_1 gene: "Pmc733"; product: "putative copper-inducible 35.6 kDa protein"; Festuca rubra putative copper-inducible 35.6 kDa protein (Pmc733) mRNA, complete cds.  
Length = 321

## HSPs:

Score = 235 (35.3 bits), Expect = 4.6e-18, P = 4.6e-18  
Identities = 59/225 (26%), Positives = 111/225 (49%)

```

Query: 647 MRELCGHLNIIYDLWSKDDHYILTSSSDGTARIWKNEINNTNTFRVLPHPSFVYTAKEH 706
      + E GH + I DLSWSK+ +L++S D T R+W ++ + +V H ++V +F+
Sbjct: 63 VHEFYGHGDAILDLSWSKNGD--LLSASMDKTVRLW--QVGRDSCLKVFSTHYVTCVQFN 119

Query: 707 PAVRELVTGTCYDSMIRIWKVEMREDSAILVRQFDVHKSFINSLCFDTGHHMYSKGDCTG 766
      P +TSC D ++RIW V LV + K + ++C+ +G +G TG
Sbjct: 120 PTNGNYFITGCDGLVRIMDVK-----CLVVDWANSKEIVTAVCYRPDGKAVAGTITG 174

Query: 767 VIVVWNTYVKINDLEHSVHHWTINKKEIKETFGIPISYLEIHPNGKRLLIHTKDSTLRI 826
      ++ +LE V ++N K ++ Y P K+L++ + D+ +RI

```

Sbjct: 175 NCRYYDASENRLEESQV---SLNGRKKSLHKRIVGFQYCPSPDP--KKLMVTSGDAQVRI 229  
 Query: 827 MDLRILVARKEVGAANYREKIHSTLTPCGTFLFAGSEGDGIVYVMN 871  
 +D +++ + G + + + + TP G + + S+D +Y+WN  
 Sbjct: 230 LGGAHVSN-YKGLQS-SSQVARSFTPDGDHIVSASDDSRIMYMN 272

Pedant information for DKFZphtes3\_ln3, frame 1

Report for DKFZphtes3\_ln3.1

[LENGTH] 1196  
 [MW] 137114.70  
 [pI] 6.79  
 [HOMOL] SWISSPROT:YKY4\_CAEEL HYPOTHETICAL 40.4 KD TRP-ASP REPEATS CONTAINING PROTEIN  
 C14B1.4 IN CHROMOSOME III. 8e-21  
 [FUNCAT] 99 unclassified proteins [S. cerevisiae, YKL121w] 2e-11  
 [FUNCAT] 04.05.01.01 general transcription activities [S. cerevisiae, YBR198c  
 TAF90 - TFIID subunit] 4e-10  
 [FUNCAT] 30.10 nuclear organization [S. cerevisiae, YBR198c TAF90 - TFIID subunit]  
 4e-10  
 [FUNCAT] 06.10 assembly of protein complexes [S. cerevisiae, YPR178w] 1e-08  
 [FUNCAT] 04.05.03 mna processing (splicing) [S. cerevisiae, YPR178w] 1e-08  
 [FUNCAT] 03.22 cell cycle control and mitosis [S. cerevisiae, YDR364c] 4e-08  
 [FUNCAT] 03.16 dna synthesis and replication [S. cerevisiae, YDR364c] 4e-08  
 [FUNCAT] 08.07 vesicular transport (golgi network, etc.) [S. cerevisiae, YDL145c]  
 9e-08  
 [FUNCAT] 30.09 organization of intracellular transport vesicles [S. cerevisiae,  
 YDL145c] 9e-08  
 [FUNCAT] 04.05.01.04 transcriptional control [S. cerevisiae, YCR084c] 2e-07  
 [FUNCAT] 10.99 other signal-transduction activities [S. cerevisiae, YHL002w] 7e-07  
 [FUNCAT] 98 classification not yet clear-cut [S. cerevisiae, YFR024c-a] 2e-06  
 [FUNCAT] 02.16 fermentation [S. cerevisiae, YMR116c] 4e-06  
 [FUNCAT] 30.03 organization of cytoplasm [S. cerevisiae, YMR116c] 4e-06  
 [FUNCAT] 05.04 translation (initiation, elongation and termination) [S. cerevisiae,  
 YMR116c] 4e-06  
 [FUNCAT] 03.10 sporulation and germination [S. cerevisiae, YFL009w] 4e-05  
 [FUNCAT] 03.04 budding, cell polarity and filament formation [S. cerevisiae, YFL009w]  
 4e-05  
 [FUNCAT] 30.04 organization of cytoskeleton [S. cerevisiae, YFL009w] 4e-05  
 [FUNCAT] 03.01 cell growth [S. cerevisiae, YCR088w] 6e-05  
 [FUNCAT] 03.25 cytokinesis [S. cerevisiae, YCR057c] 7e-05  
 [BLOCKS] BL00024H  
 [SCOP] dltbqd\_2.46.3.1.1 betal-subunit of the signal-transducing 3e-91  
 [SCOP] dlqfc\_2.21.2.1.9 Growth factor receptor-bound protein 2 (GRB2), N 4e-14  
 [SCOP] dlrmk\_2.21.2.1.8 (1-64) c-src tyrosine kinase [human (Hom 5e-15  
 [SCOP] dladd5l\_2.21.2.1.7 (1-63) Hemopoietic cell kinase Hck [human (Hom 3e-15  
 [SCOP] dlckal\_2.21.2.1.16 (1-54) p56-lck tyrosine kinase, SH3 domain [huma 1e-13  
 [SCOP] dlqwea\_2.21.2.1.15 Src kinase, SH3 domain [Avian sarcoma virus 2e-15  
 [SCOP] dlshg\_2.21.2.1.6 alpha-Spectrin, SH3 domain [chicken (Gallu 2e-13  
 [SCOP] dlprmc\_2.21.2.1.13 Src kinase, SH3 domain [chicken (Gallus gallus) 2e-15  
 [SCOP] dlhsq\_2.21.2.1.12 Phospholipase C, SH3 domain [human (Hom 2e-13  
 [SCOP] dlboa\_2.21.2.1.3 Abl tyrosine kinase, SH3 domain [Mouse (Mu 3e-13  
 [SCOP] dlfn\_2.21.2.1.2 Fyn, SH3 domain [human (Homo sapiens) 2e-15  
 [SCOP] dlsema\_2.21.2.1.11 Growth factor receptor-bound protein 2 (GRB2), N 1e-13  
 [SCOP] dlgbqa\_2.21.2.1.10 Growth factor receptor-bound protein 2 (GRB2), N 3e-16  
 [SCOP] dlckaa\_2.21.2.1.1 C-Crk, N-terminal SH3 domain [mouse (Mu 3e-15  
 [EC] 3.1.4.3 Phospholipase C 2e-07  
 [EC] 3.1.4.11 1-Phosphatidylinositol-4,5-bisphosphate phosphodiesterase 7e-07  
 [EC] 3.6.1.32 Myosin ATPase 7e-07  
 [EC] 2.7.1.112 Protein-tyrosine kinase 8e-06  
 [PIRKEW] nucleus 2e-08  
 [PIRKEW] phosphotransferase 8e-06  
 [PIRKEW] plasma 4e-07  
 [PIRKEW] duplication 4e-07  
 [PIRKEW] phosphoric diester hydrolase 2e-07  
 [PIRKEW] tandem repeat 7e-07  
 [PIRKEW] hormone 4e-07  
 [PIRKEW] transmembrane protein 2e-06  
 [PIRKEW] stomach 4e-07  
 [PIRKEW] actin binding 7e-07  
 [PIRKEW] ATP 7e-07  
 [PIRKEW] phosphoprotein 7e-07  
 [PIRKEW] signal transduction 7e-09  
 [PIRKEW] heterotrimer 7e-09  
 [PIRKEW] P-loop 7e-07  
 [PIRKEW] hydrolase 7e-07  
 [PIRKEW] transcription regulation 5e-06  
 [PIRKEW] GTP binding 7e-09

(SUPFAM) 1-phosphatidylinositol-4,5-bisphosphate phosphodiesterase II 2e-07  
 (SUPFAM) SH3 homology 2e-07  
 (SUPFAM) SH2 homology 2e-07  
 (SUPFAM) protozoan myosin heavy chain IB 7e-07  
 (SUPFAM) myosin motor domain homology 7e-07  
 (SUPFAM) pleckstrin repeat homology 2e-07  
 (SUPFAM) protein-tyrosine kinase src 8e-06  
 (SUPFAM) WD repeat homology 3e-12  
 (SUPFAM) 1-phosphatidylinositol-4,5-bisphosphate phosphodiesterase domain Y homology 2e-07  
 (SUPFAM) protein kinase homology 8e-06  
 (SUPFAM) 1-phosphatidylinositol-4,5-bisphosphate phosphodiesterase domain X homology 2e-07  
 (SUPFAM) GTP-binding regulatory protein beta chain 7e-09  
 (SUPFAM) yeast coatamer complex alpha chain 4e-07  
 (PROSITE) RGD 1  
 (PROSITE) MYRISTYL 6  
 (PROSITE) AMIDATION 2  
 (PROSITE) CAMP\_PHOSPHO\_SITE 4  
 (PROSITE) CK2\_PHOSPHO\_SITE 25  
 (PROSITE) TYR\_PHOSPHO\_SITE 4  
 (PROSITE) PKC\_PHOSPHO\_SITE 19  
 (PROSITE) ASN\_GLYCOSYLATION 6  
 (PFAM) Src homology domain 3  
 (PFAM) WD domain, G-beta repeats  
 (KW) Irregular  
 (KW) 3D  
 (KW) LOW COMPLEXITY 5.77 %  
 (KW) COILED\_COIL 2.42 %

SEQ MPTAESEAKVKTQVFEKLLKTHSDLMREKKKKKKLVRSEENISPDITIRSNLHYMKETT  
 SEG .....XXXXXXXXX.....  
 COILS .....CCCCCCCCCCCCCCCCCCCCCCCCCCCC.....  
 lgotB .....

SEQ SDDPDTIRSNLPHIKETTSDDVSAANTNNLKKSTRVTNKNKLRLNTQLATENPNGDASVEED  
 SEG .....  
 COILS .....  
 lgotB .....

SEQ RQGKPNKKVKTVPOLTTQDLKPETPENKVDSTHQRTHTKPPQGVDPKSEKANEGREET  
 SEG .....XXX  
 COILS .....  
 lgotB .....

SEQ DLEDEELMQAYQCHVTEENAKEIKRKIRKKLKEQLTYFPSDTLFDHDKLSSEKRRKKKKE  
 SEG .....XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX  
 COILS .....  
 lgotB .....

SEQ VPFVSKAETSTLTISGDTVEGEQKKFSSVRSVSSDSHQDDEISSMEQSTEDSMQDDTKPK  
 SEG .....XXXXXXXXX.....XXX  
 COILS .....  
 lgotB .....

SEQ PKKTKKKTRAVADNEDVDGQVHEITSRDSVPYPKCLDDDLVLGVYIHRDLRLKSDFM  
 SEG .....XXXXXXXXX.....  
 COILS .....  
 lgotB .....

SEQ ISHPMVKIHVVDEHTGQYVKKDDSGRPVSSYYEKENVDYILPIMTQPYDFKQLKSRLPEW  
 SEG .....  
 COILS .....  
 lgotB .....

SEQ EEQIVFNENFPYLLRGSDESPKVLFFEILDFLSVDEIKNNSEVQNEQCGFRKIAMAFK  
 SEG .....  
 COILS .....  
 lgotB .....

SEQ LLGANGNANINSKLRLQLYYPPTKPRSPLSVVEAFEMWSKCPRNHYPTLYTVTRGLKVP  
 SEG .....  
 COILS .....  
 lgotB .....

SEQ DCIKPSYRSMALQEEKGPVHCERHHESSSVOTEPGLEESKEVIKWRKLPQACRIPNK  
 SEG .....  
 COILS .....  
 lgotB .....

```

SEQ      HLFSLNAGERGCFCLDFSHNGRILAAACASRDGYPIILYEIPSGRFMRELCGHLNIIYDL
SEG      .....
COILS    .....
lgotB    .....CEEEEECCCCCEEE

SEQ      SWSKDDHYILTSSSDGTARIWKNEINNTNTRVLPHPSFVYTAKFHPAVRELVTTCYDS
SEG      .....
COILS    .....
lgotB    EETTTTTEEEETTTEEEETT--TTCEEEETTTCCEEEETT--TCCEEEETT

SEQ      MIRIWKVEMREDSAILVRQFDVHKSFINSLCFDTEGHHMYSGDCTGVIVVWNTYVKINDL
SEG      .....
COILS    .....
lgotB    EEEEEETTTTBTTEEEEEECCCE-EEEEETTEEEETTTEEEEEE.....

SEQ      EHSVHHWTINKEIKETEFKGIPIISYLEIHPNGKRLLIHTKDSTLRIMDLRILVARKEVGA
SEG      .....
COILS    .....
lgotB    .....

SEQ      ANYREKIHSTLTPCGTFLFAGSEGDIVVWNPETGEQVAMYSOLFPKSPIRDISYHPFEN
SEG      .....
COILS    .....
lgotB    .....

SEQ      MVAFCAFQGNPILLYIDFVHAQQAEMFKRYNGTFPLPGIHQSQDALCTCPKLPHQGS
SEG      .....
COILS    .....
lgotB    .....

SEQ      FQIDFVHTESSTKMQLVKORLETVTEVIRSCAAKVNKNLSFTSPPAVSSQSKLKQSN
SEG      .....
COILS    .....
lgotB    .....

SEQ      MLTAQEILHQFGFTQTGIISIERKPCNHQVDTAPTVALYDYTANRSDTLTHRGDIIRV
SEG      .....
COILS    .....
lgotB    .....

SEQ      FFKDNEDWYGSIGKGQEGYFPANHVASETLYQELPPEIKERSPPLSPEEKTKIEKSPAP
SEG      .....
COILS    .....
lgotB    .....

SEQ      QKQSINKNKSQDFRLGSESMTHSEMRKEQSHEDQGHIMDTMRKNKQAGRKVTLIE
SEG      .....
COILS    .....
lgotB    .....

```

## Prosites for DKFzphes3\_in3.1

PS00001	460->464	ASN_GLYCOSYLATION	PDOC00001
PS00001	686->690	ASN_GLYCOSYLATION	PDOC00001
PS00001	934->938	ASN_GLYCOSYLATION	PDOC00001
PS00001	1000->1004	ASN_GLYCOSYLATION	PDOC00001
PS00001	1065->1069	ASN_GLYCOSYLATION	PDOC00001
PS00001	1148->1152	ASN_GLYCOSYLATION	PDOC00001
PS00004	91->95	CAMP_PHOSPHO_SITE	PDOC00004
PS00004	264->268	CAMP_PHOSPHO_SITE	PDOC00004
PS00004	305->309	CAMP_PHOSPHO_SITE	PDOC00004
PS00004	1190->1194	CAMP_PHOSPHO_SITE	PDOC00004
PS00005	48->51	PKC_PHOSPHO_SITE	PDOC00005
PS00005	66->69	PKC_PHOSPHO_SITE	PDOC00005
PS00005	93->96	PKC_PHOSPHO_SITE	PDOC00005
PS00005	170->173	PKC_PHOSPHO_SITE	PDOC00005
PS00005	232->235	PKC_PHOSPHO_SITE	PDOC00005
PS00005	268->271	PKC_PHOSPHO_SITE	PDOC00005
PS00005	304->307	PKC_PHOSPHO_SITE	PDOC00005
PS00005	327->330	PKC_PHOSPHO_SITE	PDOC00005
PS00005	352->355	PKC_PHOSPHO_SITE	PDOC00005
PS00005	384->387	PKC_PHOSPHO_SITE	PDOC00005
PS00005	440->443	PKC_PHOSPHO_SITE	PDOC00005
PS00005	533->536	PKC_PHOSPHO_SITE	PDOC00005
PS00005	546->549	PKC_PHOSPHO_SITE	PDOC00005
PS00005	643->646	PKC_PHOSPHO_SITE	PDOC00005
PS00005	677->680	PKC_PHOSPHO_SITE	PDOC00005
PS00005	690->693	PKC_PHOSPHO_SITE	PDOC00005
PS00005	702->705	PKC_PHOSPHO_SITE	PDOC00005



PS00005	823->826	PKC_PHOSPHO_SITE	PDOC00005
PS00005	973->976	PKC_PHOSPHO_SITE	PDOC00005
PS00006	22->26	CK2_PHOSPHO_SITE	PDOC00006
PS00006	59->63	CK2_PHOSPHO_SITE	PDOC00006
PS00006	77->81	CK2_PHOSPHO_SITE	PDOC00006
PS00006	116->120	CK2_PHOSPHO_SITE	PDOC00006
PS00006	137->141	CK2_PHOSPHO_SITE	PDOC00006
PS00006	180->184	CK2_PHOSPHO_SITE	PDOC00006
PS00006	245->249	CK2_PHOSPHO_SITE	PDOC00006
PS00006	276->280	CK2_PHOSPHO_SITE	PDOC00006
PS00006	283->287	CK2_PHOSPHO_SITE	PDOC00006
PS00006	288->292	CK2_PHOSPHO_SITE	PDOC00006
PS00006	292->296	CK2_PHOSPHO_SITE	PDOC00006
PS00006	327->331	CK2_PHOSPHO_SITE	PDOC00006
PS00006	390->394	CK2_PHOSPHO_SITE	PDOC00006
PS00006	454->458	CK2_PHOSPHO_SITE	PDOC00006
PS00006	510->514	CK2_PHOSPHO_SITE	PDOC00006
PS00006	570->574	CK2_PHOSPHO_SITE	PDOC00006
PS00006	663->667	CK2_PHOSPHO_SITE	PDOC00006
PS00006	672->676	CK2_PHOSPHO_SITE	PDOC00006
PS00006	804->808	CK2_PHOSPHO_SITE	PDOC00006
PS00006	985->989	CK2_PHOSPHO_SITE	PDOC00006
PS00006	1023->1027	CK2_PHOSPHO_SITE	PDOC00006
PS00006	1127->1131	CK2_PHOSPHO_SITE	PDOC00006
PS00006	1132->1136	CK2_PHOSPHO_SITE	PDOC00006
PS00006	1161->1165	CK2_PHOSPHO_SITE	PDOC00006
PS00006	1170->1174	CK2_PHOSPHO_SITE	PDOC00006
PS00007	1083->1091	TYR_PHOSPHO_SITE	PDOC00007
PS00007	211->219	TYR_PHOSPHO_SITE	PDOC00007
PS00007	1083->1091	TYR_PHOSPHO_SITE	PDOC00007
PS00007	210->219	TYR_PHOSPHO_SITE	PDOC00007
PS00008	483->489	MYRISTYL	PDOC00008
PS00008	577->583	MYRISTYL	PDOC00008
PS00008	716->722	MYRISTYL	PDOC00008
PS00008	800->806	MYRISTYL	PDOC00008
PS00008	861->867	MYRISTYL	PDOC00008
PS00008	941->947	MYRISTYL	PDOC00008
PS00009	811->815	AMIDATION	PDOC00009
PS00009	1188->1192	AMIDATION	PDOC00009
PS00016	1074->1077	RGD	PDOC00016

## Pfam for DKF2phtes3\_1n3.1

HMM_NAME	WD domain, G-beta repeats		
HMM	*MrGHnnWVWCvAFSPDGzWFIvSGSWDgTCRLWD*		
Query	650	LCGHLNIIYDLWSKDDHY-ILTSDDGTARIWK	682
HMM_NAME	Src homology domain 3		
HMM	*pyVIALYDYqAqdpDELSFkEGDIIiIEdeDD.WWrgRnnnTNGQEGW		
	P+V+ALYDY+A+++DEL++ +GDII + +++ WW+G GQEG+		
Query	1054	PTVVALYDYTANRSDELTIHRGDIIIRVFFKDNEWWYGSIGK--GQEGY	1100
HMM	IPSNYVEPI*		
	+P+N V+ +		
Query	1101	FPANHVAE	1109

DKFZphtes3\_20c21  
-----

group: testes derived

DKFZphtes3\_20c21 encodes a novel 708 amino acid protein without similarity to known proteins.

No informative BLAST results; No predictive prosite, pfam or SCOP motif.

The new protein can find application in studying the expression profile of testis-specific genes.

unknown

Sequenced by MediGenomix

Locus: /map="22q11.2-12.2"

Insert length: 3997 bp

Poly A stretch at pos. 3877, polyadenylation signal at pos. 3853

```
1 GGTAGGCGGG GCGGCGCGTG ACCTAAGGCC TCTCTGCCGC GCGGCGAGGT
51 ACGGGGCGAGA AGTCGCGAGGT ACCCAGCTGC TGCCACGTTT TCTGGTCCAG
101 AGTCCCGAAC CCCGAGCACT GGGATGGCTG GCTACTCCGA GCCAAGCCAC
151 TGATGTTTGA ACTGGAAACT TCAGAACGTT TAATAAGACT CTTCAGGATG
201 GGTITGAACCT AGACAACCTA GAAATTTCTT TAGAACACCA GCTCTAGCAT
251 GCATCTCCCA CTTTTGGCTT TCCTGGAGAG GAGCTTGAAG AGGTGTCTCT
301 GCAGACAGCC ACAGTGATAC TCAGGAACCC AGAGGAATGG ATTTGACTTT
351 TCTGTAGGA TTCTTTGTTA TAGTTTCTCC CTGAGTGTGA AGAGGCATGG
401 AAATATACAT GAAACTGAAG AACCTGCAAG GAAGGGAAGT GGAACTTTCC
451 ATGCTGAGTG AAAACTAACC AAGTGGCAGT TGTGACTGAA AACACTGAAA
501 CCTACCACGT CCAGATTAC TGGATTGGGG GATAGAGGAA CGGTACACGC
551 TAGGGAGAAA GAAGTGATAC CGGAAAAGAA AACCTAAATG AAGAGAATGA
601 GGATGACTGC ACAGTAGATG GCCACCTCTA CTCCACAGA GGCAAAGTCA
651 GCCTCGTGGT GGAATTATTT TTTCTTTAT GATGGTTCCA AGGTAAAGGA
701 AGAAGGCGAT CCAACAAGAG CTGGCATTGG TTACTTTTAT CTTTCCAGA
751 CCCTGCTAGA CCAACAGGAG TTGCTTTTGG GACAGATTGC TGGAGTTGTC
801 CGCTGTGTTT CTGACATTTC TGACTCTCCT CTTACTCTTG TTGCTGTGAG
851 AAAACTGAAG TTGGCATAAA AAGTTGATGG AGATTACCTT TGGGTCTGGG
901 GCTGTGCTGT GGAGCTCCCT GATGTACGCT GCAGCGGTTT TCTGGATCAG
951 CTAGTTGGAT TCTTTAATTT TTACAATGGA CTTGTTTCCC TAGCTTATGA
1001 GAAGTGTCTC CAGGAAGAAG TGAGCACGGA GTGGGACACC TTCTCGAGC
1051 AAATTCTGAA AAACACCACT GATCTGCATA AGATTTTCAA TTCCCTCTGG
1101 AACTTGGACC AAATAAAGT GGAGCCCCTG TTGTTGCTGA AGGCAGCCCG
1151 CATTCTGCAG ACCTGCCAGC GCTCGCTCA CATTCTCGCT GGCTGCATCC
1201 TCTATAAAGG ACTGATTGTC AGCACCCCA TCCCGCCCTC CCTCACCGCC
1251 AAGGTCCCTG TCCACCGAAC AGCACCTCAG GAGCAGAGAC TCCCTACGGG
1301 AGGGGATGCC CCGCAGGAAC ATGGAGCGGC ATTGCCCTCC AATGTCCAGA
1351 TTATCCCTGT TTTTGTGACC AAAGAGGAAG CCATTAGTCT CCACGAGTTC
1401 CCGGTGGAAC AGATGACAAG GTCTCTAGCA TCTCCAGCAG GACTCCAGGA
1451 TGGTTCAGCC CAGCACCATC CAAAGGGTGG GAGCACATCT GCCCTGAAAG
1501 AAAACGCCAC TGGCCATGTT GAATCCATGG CTTGGACCAC CCCAGATCCC
1551 ACATCCCTCT ACGAAGCTTG TCCAGATGGC AGGAAGGAGA ACGGATGCTT
1601 GTCTGGCCAT GATCTGGAGA GCATCAGGCC CGCAGGACAT CACACTCTG
1651 CCAGGGGTGA GGTCTCTGGC CTCAGCTCCT CCTTGGGAAA GGAAGTAGTC
1701 TTTCTCCAGG AAGAACTCGA CTTGTCTGAA ATCCACATTC CAGAGGCTCA
1751 GGAAGTGGAA ATGGCCTCAG GTCAATTTGC TTCTCTACAT GTGCCTGTTT
1801 CAGATGGCAG GGCTCCTTAC TGCAAGGCAT CTCTCAGCGC CTCACAGCAG
1851 CTGGAACCCA CGCTCTCTGA GCACAGAGC ATCAGCAGCT TGCGCCCTCC
1901 CTCTGCTCCT GAGATGCTGA CCCAGCATGG AGCCCAAGAG CAGGTGGAAG
1951 ACCATCTCTG CCATAGCAGC CAAGCCCCCA TTCCAGAGC AGACCTCTCT
2001 CCCAGAAGGA CCGCAGGCC CTTGTTATTG CCTCGCTTAG ATCCAGGACA
2051 GAGAGAAAC AAGCTTCCCA CGGGGGAACA AGGCCTGGAT GAGGATGTTG
2101 ATGGGGTCTG TGAAGCCAC GCAGCCCCTG GTCTGGAATG CAGTTGAGGC
2151 TCAGCAAACT GTCAGGGTGC TGGCCCTCTC GCAGATGGAA TCAGTCCAG
2201 GCTGACACCA GCAGAGTCTT GCATGGGGCT CGTGAGGATG AATCTCTACA
2251 CTCACTGCGT CAAAGGGCTG ATGCTGTCCC TGCTGGCTGA GGAGCCGCTG
2301 CTGGGAGACA GCGCAGCCAT AGAGGAAGTG TACCACAGCA GCCTGGCTTC
2351 ACTGAATGGG CTGGAAGTCC ACCTGAAAGA GAGCTGCCCC AGGATGAGG
2401 CAGCTCTCAC GAGCAGCAGC TACAACCTTA CATATTACGA CGCATTCAG
2451 AGCTTGCTGA TGGCAAACTT CCGCAGGCTG GCCACCCGCG ATGATCGCCG
2501 CTCTCTCCAG GCGCTCAGCC TGATGCAATG GCAATTTGCC CAGCTGCCCC
2551 CGCTTTATGA AATGACTGTC AGAATGCCT CCACGGCTGT GTAGCCTGTG
2601 TCCACCCCA TCCAGGAGAC ATATTTCCAG CAGCTGGCAG CTCAGCAGC
2651 GAGCTCCGCG TTCCCAJACC CTCAGGATGG CGCTTCAGC CTCTCCGGCA
2701 AAGCAAGCA GAAGCTGCTG AAGCAGGGGG TGAATTTGCT CTGAATGCA
2751 CCCAGGAGGT GACTGGGAAG GAGAAAACCA GCAAAGGAAG CTCTGCTTTT
2801 TATAATTGAA AAGGCCCTCT TATTTTATTT TTCTTGAATA CATTCCCTTT
```

```
2851 TTTAGGAACC AAATGATATT TGAGTTTTTG TTATTCCTTT TGCAGATTGG
2901 GATGTGTTTT GGGGGCAGGG CTTAGTTCTT CAGGTGGGCA GACCCAGAGC
2951 ACTTGATAAA GAACGTGATT TAATCGGTAG TGTGGGGGCC GGGACGGGCT
3001 TGGCTCCCTC TCTGCCATAC TGAGCCTGAG GTATTTCATA TCTCCTGCTG
3051 TTCCATCCCA GCTTGAATTG GTGCCACAAG CTTCCAAGTT GGCATTTTTT
3101 CTAGAACCCTG ATCGTCCACT AGCCCAGAGT GTGTGTGTC AACCCCCACA
3151 CCAGGTGGTG GTAGGCGGTG TGACTGCACA GCGAGGTGCC GGATCTGTGA
3201 GCAGGCCGAC TCCACTCCCA CGCCGAGGT AGGTTTCTCC AGTGGGCTCT
3251 TGCTGGGAGG TCCGGATCGT TCCTGCAGGG AAGCGGCAGC ACACGGAGAC
3301 CACTTGGTTG AATTCTGTG GAACTCTACT CAAATCTAGG GGCCTCTTCT
3351 TTGAGCCAC AATGGGGGCA AGCCTTAATA ATATGGAAGG GAGTTTGGGC
3401 TTTAGAGATC CCTTTATAAA AGCTCTGGGG GCTGAGCCCT GAGAATTCAG
3451 TGACAACAGG ACCAACCTGC GTGGCCTTTG ACTACAAGTG GGCCTGTCAG
3501 CTGGTTCTCT TCGAGCGAGT GTCCCTAAAT AGGAGTTTAC AAGATGCTTG
3551 GGGGTAAAG CACTGTGCTT TTCAGTGGTG CTTGCGTAA AGGAGCGGAC
3601 ACTCAGCTGT GTGTCTCTGG GCTTGTCTGG TACTTGAAC CTCAGTTCTA
3651 TTACGTTATA CTCAGACATT TTTTGTGACG TATGAGACAG ACTGCAGGAT
3701 GAAATATTT GTCAAAATCT TAACGAATG TTTACTGAA GTACTTGAGA
3751 TTCCATTGGA GAGTTGTATT GTTAATAAT TCATGTCAGT GAACGTGAT
3801 CTGATGTTTA TGATATGGTG TCTTTTCTT GAAACAAGCT TCCAGGGCT
3851 AGAAATAAAA TAGCCAAAAA ATGCTGGAAA AAAAAAAAAA AAAAAAAAAA
3901 AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA
3951 AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAA
```

## BLAST Results

Entry HS1048E9 from database EMBLNEW:  
Human DNA sequence from clone 1048E9 on chromosome 22q11.2-12.2  
Contains pseudogene similar to ribosomal protein S3A and part of a gene  
similar to C.elegans protein CE02118, ESTs, STS, GSS.  
Score = 6540, P = 0.0e+00, identities = 1308/1308  
-14 exons

## Medline entries

No Medline entry

## Peptide information for frame 3

ORF from 618 bp to 2741 bp; peptide length: 708  
Category: putative protein  
Classification: no clue

```
1 MATSTSTEAK SASWNNYFFL YDGSKVKEEG DPTRAGICYF YPSOTLLDQO
51 ELLCQGIAGV VRCVSDISDS PPTLVRLRKL KFAIKVDG DY LHWLGC AVEL
101 PDVSKRFLD QLVGFNFYN GPVSLAYENC SQEELSTEW TFIEQILKNT
151 SDLHKIFNSL WNLQOTKVEP LLLKAAARIL QTCQRS PHIL AGCILYKGLI
201 VSTQLPPLSL AKVLLHRTAP OEORLPTGGD APQEHGAALP PNVQIIPV FV
251 TKEEAILSHE FPVEQMTRSL ASPAGLQDGS AQHHPKGGST SALKENATGH
301 VESMAWITPD PTPDEACPD GRKENGCLSG HDLESIRPAG LHNSARGEVL
351 GLSSSLGKEL VFLQEELDLS EIHIPEAQEV EMASGHFAFL HVPVPOGRAP
401 YCKASLSASS SLEPTPPEDT AISSLRPPSA PEMLTQHGAQ EQVEDHPGHS
451 SQAPIPRADP LPRRTRRPLL LPRLDPGQRG NKLPTGEQGL DEDVDGVCE S
501 HAAPGLECSS GSANCQGAGP SADGISSRLT PAESCMGLVR MNLYTHCVKG
551 LMLSLLAEPP LLGDSAAIEE VYHSSLASLN GLEVHLKETL PRDEAAS TSS
601 TYNFTYDRI QSLLMANLPQ VATPHDRRFL QAVSLMHSEF AQLPALYEM T
651 VRNASTAVYA CCNPIQETYP QQLAPAARSS GFNPNQDQAF SLSGKAKQKL
701 LKHGVNLL
```

## BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphtes3\_20c21, frame 3

No Alert BLASTP hits found

Pedant information for DKFZphtes3\_20c21, frame 3

Report for DKFZphtes3\_20c21.3

(LENGTH) 708  
(MW) 76900.23  
(pI) 5.30  
(KW) Alpha\_Beta  
(KW) LOW\_COMPLEXITY 6.36 %

SEQ MATSTSTEAKSASWNYFFLYDGSKVKEGDPTRAGICYFYPSQTLDDQQLLCGQIAGV  
SEG .XXXXXXXXXXXXX.....  
PRD ccc

SEQ VRCVSDISDSPPTLVRLRLKLFKIVDGYLWVLGCAVELPDVSKRFLDQLVGFFNFYN  
SEG .....  
PRD eeeeecc

SEQ GPVSLAYENCSEELSTEWDTFIEQILKNTSDLHKIFNSLWNLDTKVEPLLLKKAARIL  
SEG .....  
PRD ccc

SEQ QTCQRSPHILAGCILYKGLIVSTQLPPSLTAKVLLHRTAPQEQRLPTGGDAPQEHGAALP  
SEG .....  
PRD hhhcc

SEQ PNQIIPVFVTKKEATSLHEFPVEQMTSLASAPAGLDGSAQHHPKGGSTSAKENATGH  
SEG .....  
PRD ccc

SEQ VESMAWTTDPDTPSPDEACPDGRKENGCLSGHDLESIRPAGLHNSARCEVLGLSSSLGKEL  
SEG .....  
PRD ccc

SEQ VFLQEELDLSEIHIEAQEVEMASGHFAFLHVPVDPGRAPYCKASLSASSSLEPTPPEDT  
SEG .....  
PRD hhhhhhhcc

SEQ AISSLRPPSAPEMLTQHGAQEQVEDHPGHSSQAPIPRADPLPRRTLRPLLPRLDPGQRG  
SEG .....  
PRD ccc

SEQ NKLPTEGQGLDEDVDGVCESHAAPGLECSSGSANCQAGPSADGSSRLTPAESCHGLVR  
SEG .....  
PRD ccc

SEQ MNLYTHCVKGLMLSLLAEEPLLGDSAAIEEVYHSSLASLNGLEVHLKETLPDEAASTSS  
SEG .....  
PRD ccc

SEQ TYNFTYYDRIQSLLMANLPQVATPHDRRLQAVSLMHSEFAQLPALYEMTVRNASTAVYA  
SEG .....  
PRD ccc

SEQ CCNPIQETTFQQLAPAAARSSGFNPQDGAFLSGKAKQKLLKHGVNLL  
SEG .....  
PRD ccc

(No Prosite data available for DKF2phtes3\_20c21.3)

(No Pfam data available for DKF2phtes3\_20c21.3)

DKFZphtes3\_20k2

-----

group: signal transduction

DKFZphtes3\_20k2 encodes a novel 839 amino acid protein with strong similarity to rat vanilloid receptor subtype 1.

VR1 seems to play an important role in the activation and sensitization of nociceptors. It is the receptor for e.g. capsaicin, a selective activator of nociceptors, a natural product of capsicum peppers. The novel protein is the human orthologue of rat VR1.

The new protein can find application as a target for the development of new nociception-modulating drugs.

strong similarity to rat vanilloid receptor subtype 1

Sequenced by MediGenomix

Locus: unknown

Insert length: 4187 bp

Poly A stretch at pos. 4154, polyadenylation signal at pos. 4135

```
1 GGCTCAGGCA GGCCTGGCCC AGAGTCACGC TGGCAACCCAG GAGTTTGGGA
51 AGCAGTCGTA TTCTCTCTCT CTCTCTCTCT CTCTCAGTAT CCATGACAGT
101 GTGATGGAGA GTCTCTGGCG TGCCATCTGG GATGCAAAAC GTCCCTGTGT
151 CCCCACGCTC CAGGCCGTAG ATGCTCCCCG CGGTCAGTC ACTTAGTCGT
201 CAGATCGGCC GTCTCGTAT CACAGTGCCT GTGTTCAAGT TGCACACTGG
251 GCCACAGAGG ATCCAGCAAG GATGAAGAAA TGGAGCAGCA CAGACTTGGG
301 GGCAGCTGGC GACCACTCC AAAAGGACAC CTGCCAGAC CCCCTGGATG
351 GAGACCCCTAA CTCACAGGCC CCTCCAGCCA AGCCCCAGCT CTCACAGGCC
401 AAGAGCCGCA CCGGCTCTTT TGGGAAGGGT GACTCGGAGT AGGCTTTCCC
451 GGTGGATTGC CCTCACGAGG AAGGTGAGCT GGAATCCTGC CCGACCATCA
501 CAGTCAGCCC TGTTATCACC ATCCAGAGGC CAGGAGACGG CCCACCCGGT
551 GCCAGGCTGC TGTCCTCAGG CTCTGTCTGC GCCAGCACCG AGAAGACCCT
601 CAGGCTCTAT GATCGCAGGA GTATCTTTGA AGCCGTTGCT CAGAATAACT
651 GCCAGATCTT GGAGAGCCCTG CTGCTCTTCC TGCAGAGAG CAAGAAGCAC
701 CTCACAGACA ACAGGTTCAA AGACCTGAGC ACAGGGAAGA CTTCTCTGCT
751 GAAAGCCATG CTCACCTGCG ATGACGGACA GAACACACCC ATCCCCCTGC
801 TCCTGGAGAT CGCGCGGCAA ACGGACAGCC TGAAGGAGCT TGTCAACGCC
851 AGCTACACGG ACAGTACTTA CAAGGGCCAG ACAGCACTGC ACATCGCCAT
901 CGAGAGACGC AACATGGCCC TGGTGACCTT CCTGGTGAG AACGGAGCAG
951 ACGTCCAGGC TGGCGCCCAT GGGGACTTCT TTAAGAAAAC CAAAGGGCGG
1001 CCTGGATTCT ACTTCGGTGA ACTGCCCTTG TCCCTGGCCG CGTGACCAA
1051 CCAGCTGGGC ATCGTGAAGT TCCTGCTGCA GAACTCCTGG CAGACGGCCG
1101 ACATCAGCGC CAGGGAATCG GTGGGCAACA CGGTGCTGCA CGCCCTGGTG
1151 GAGGTGGCCG ACAACACGGC CGACAACACG AAGTTTGTGA CGAGCATGTA
1201 CAATGAGATT CTGATCCTGG GGCCCAACT GCACCCGAGG CTGAAGCTGG
1251 AGGAGCTCAC CAACAAGAAG GGAATGACGC CGCTGGCTCT GGCAGCTGGG
1301 ACCGGGAAGA TCGGGGCTTT GGCCTATATT CTCCAGCGGG AGATCCAGGA
1351 GCCCGAGTGC AGGCACCTGT CCAGGAAGTT CACCGAGTGG GCCTACGGGC
1401 CCGTGCACTC CTGCGCTGAC GACCTGTCTT GCATCGACAC CTGCGAGGAG
1451 AACTCGGTGC TGGAGGTGAT CGCCTACAGC AGCAGCGAGA CCCCTAATCG
1501 CCACGACATG CTCTTGGTGG AGCCGCTGAA CCGACTCCTG CAGGACAAGT
1551 GGGACAGATT CGTCAAGCGC ATCTTCTACT TCAACTTCTT GGTCTACTGC
1601 CTGTACATGA TCATCTTCC CATTGGCTGC TACTACAGGC CCGTGGATGG
1651 CTTGCTTCCC TTTAAGATGG AAAAAATTGG AGACTATTTC CGAGTTACTG
1701 GAGAGATCCT GTCTGTGTTA GGAGGAGTCT ACTTCTTTTT CCGAGGGATT
1751 CAGTATTTCC TGCAGAGGCG GCGCTCGATG AAGACCCCTG TTGTGGACAG
1801 CTACAGTAGA ATGCTTTTCT TTCTGCACTG ACTGTTCTAG CTGGCCACCG
1851 TGGTCTGTGA CTTCAGCCAC CTCAGGAGAT ATGTGGCTTC CATGGTATTC
1901 TCCTTGGCCT TGGGCTGGAC CAACATGCTC TACTACACCC GCGGTTTCCA
1951 GCAGATGGGC ATCTATGCCG TCATGATAGA GAAGATGATC CTGAGAGACC
2001 TGTGCCGTTT CATGTTGTTC TACATCGTCT TCTTGTTCGG GTTTTCCACA
2051 GCGGTGGTGA CGCTGATTGA AGACGGGAAG AATGACTCCC TGCCGTCTGA
2101 GTCCACGTGC CACAGGTGGC GGGGGCCTGC CTGCAGGCCC CCCGATAGCT
2151 CCTACAACAG CCTGTACTCC ACCTGCCTGG AGCTGTTCAA GTTCAACATC
2201 GGCATGGGCG ACCTGGAGTT CACTGAGAAC TATGACTTCA AGGCTGTCTT
2251 CATCATCCTG CTGCTGGCCT ATGTAATCTT CACCTACATC CTCCTGTCTA
2301 ACATGCTCAT CGCCCTCATG GGTGAGACTG TCAACAAGAT CGCACAGGAG
2351 ACAGAAGACA TCTGGAAGCT GCAGAGAGCC ATCCATATCC TGGACACGGA
2401 GAAGAGCTTC CTTAAGTGCA TGAGGAAGGC CTTCGCTCA GGCAGCTGTC
2451 TGCAGGTGGG GTACACACCT GATGGCAAGG ACGACTACCG GTGGTGCTTC
2501 AGGGTGGACG AGGTGAACCT GACCACTGAG AACCCAACG TGGGATCAT
2551 CAACGAAGAC CCGGCAACT GTGAGGCGGT CAAGCGCACC CTGAGCTTCT
2601 CCCTGCGGTC AAGCAGAGTT TCAGGCAGAC ACTGGAAGAA CTTTGCCCTC
2651 GTCCCCCTTT TAAGAGAGGC AAGTGCTCGA GATAGGCAGT CTGCTCAGCC
```

```
2701 CGAGGAAGTT TATCTGGGAC AGTTTTCAGG GTCTCTGAAG CCAGAGGACG
2751 CTGAGGTCTT CAAGAGTCCT GCCGCTTCCG GGGAGAAGTG AGGACGTCAC
2801 GCAGACAGCA CTGTCAACAC TGGGCTTAGG GAGACCCCGT TGCCACGGGG
2851 GGCTGCTGAG GGAACACAGG TGCTCTGTGA GCAGCCTGGC CTGGTCTGTG
2901 CCTGCCGAGC ATGTTCCCAA ATCTGTGCTG GACAAGCTGT GGGAGCCTTT
2951 CTGGGAAGCA TGGGAGTGA TGTACATCCA ACCGTCACGT GCCCAAGTG
3001 AATCTCTTAA CAGACTTTCA GGTTTTACT CACTTTACTA AACAGTTTGG
3051 ATGGTCAGTC TCTACTGGGA CATGTTAGGC CCTTGTTC TTTGATTTTA
3101 TTCTTTT TTTGAGACAGAA TTTCACTCT CTCACCCAGG CTGGAATGCA
3151 GTGGCACAAT TTTGGCTCCC TGCAACCTCC GCCTCCTGGA TTCCAGCAAT
3201 TCTCTGGCCT CGGCTTCCCA AGTAGCTGGG ATTACAGGCA CGTGCCACCA
3251 TGTCTGGCTA ATTTTGTGTA TTTTCTTAAT AGATATGGGG TTTCGCCATG
3301 TTGGCCAGGC TGGTCTCGAA CTCCTGACCT CAGGTGATCC GCCCACCTCG
3351 GCCTCCCAAA GTGCTGGGAT TACAGGTGTG AGCCTCCACA CCTGGCTGTT
3401 TTCTTTGATT TTATCTTTT TTTTCTTCT GTGAGACAGA GTTCACTCT
3451 TGTGCCCAGG GCTGGAGTGC AGTGGTGTGA TCTTGGCTCA CTGCAACCTC
3501 TGCTCTCCCG GTTCAAGCGA TCTTCTGCT TCAGTCTCCC AAGTAGCTTG
3551 GATTACAGCT GAGCACTACC ACCGCCGCT AATTTTGTG TTTTAAATAG
3601 AGACGGGGTT TCACCATGTT GGCAGGCTG GTCTCGAAT CTGACCTCA
3651 GGTGATCTGC CGGCTTGGC CTCCCAAAGT GCTGGGATTA CAGTCTGAG
3701 CGGCTGGCCT CGGCTTCTT TGATTTTATA TTATTAGGAG CAAAGTAAA
3751 TGAAGCCAG GAAACACCT TTGGGAACAA ACTCTCTCT TGATGAAAA
3801 TGCAGAGGCC CTTCCTCTCT GTGCGTGTG TGCTCTCTT ACCTGCCCGG
3851 GTGGTTTGGG GGTGTTGGTG TTTCTCTCT GGAGAAGATG GGGGAGGCTG
3901 TCCCACTCCC AGCTTGGCA GAATCAAGCT GTTGACAGC TGCTTCTTC
3951 ATCCTCTCTT ACGATCAATC ACAGTCTCCA GAAGATCAGC TCAATTGCTG
4001 TGCAGGTAA AACTACAGAA CCACATCCCA AAGGTACCTG GTAAGAATGT
4051 TTGAAAGATC TTCCATTCTT AGGAACCCCA GTCTGCTTC TCCGAATGG
4101 CACATGCTTC CACTCCATC ATACTGGCAT CCTCAATTA ACAGATATGT
4151 ATACATATA AAAAAAAAAA AAAAAAAAAA AAAAAA
```

## BLAST Results

No BLAST result

## Medline entries

99288727:  
Recent advances in neuropharmacology of cutaneous nociceptors.

99231880:  
A non-pungent triprenyl phenol of fungal origin, scutigerol, stimulates  
rat dorsal root ganglion  
neurons via interaction at vanilloid receptors.

## Peptide information for frame 2

ORF from 272 bp to 2788 bp; peptide length: 839  
Category: strong similarity to known protein  
Classification: Cell signaling/communication

```
1 MKKWSSTDLG AAADPLQKDT CPDPLOGOPN SRPPPAKPQL STAKSRTLRF
51 GKGDSSEAFV VDCPHEEGEL DSCPTITVSP VITIQRPGDG PTGARLLSQD
101 SVASSTEXTL RLYDRSIEE AVKQNNQOOL ESILLFLQKS KKHLDNEFK
151 DPEFGATCLL KAMLNLDGGO NTPIPLLEI ARQDLSLEL VNASYDTSYY
201 KGQALHIAI EERNMALVTL LVENGADVQA AAHGFYFKT KGRPGFYFGE
251 LPLSLAECTN QLGIVKFLQ NSQTDADISA RDSVGNVVLH ALVEVADNTA
301 DNTKFVTSY NEILILGAKL HPTLKLELT NKKGMTPLAL AAGTGKIGVL
351 AYILQREIQE PECHLSRKF TEWAYGPVHS SLYDLSCIDT CEKNSVLEVI
401 AYSSETPNR HDMLLVEPLN RLLQDKWDRF VKRIFYFNFL VYCLYMIIFT
451 MAAYRFPVDG LPPFRMEKIG DYFRVTGEIL SVLGGVYFFF RGIQYFLQRR
501 PSMKTLFVDS YSEMLFFLQS LFMLATVVLY FSHLKEYVAS MVFSLALGWT
551 NMLYYTRGFQ QMGIIYAVMIE KMILRDLCRF MFVYIVFLFG FSTAVVTLIE
601 DGKNDLPSSE STSHRWGPA CRPDSSVNS LYSTCLELFK FTIGMDLEF
651 TENYDFKAVF ILLLLAVIL TYILLAMLI ALMGETVNNI AQESKNIWL
701 QRAITLQTE KSEFLKMRKA FRSGKLLQW YTPDGKDYR WCFRVEVNN
751 TTMNTNVCII NEDPGNCEGV KRTLSFLRS SRVSGRHWKN FALVPLLEA
801 SARDRQSAQP EEVYLRQFSG SLRPDAEVF KSPAASCEK
```

## BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKF2phtes3\_20k2, frame 2

TREMBL:AF029310.1 product: "vanilloid receptor subtype 1"; Rattus norvegicus vanilloid receptor subtype 1 mRNA, complete cds., N = 1, Score = 3760, P = 0

TREMBL:NEW:AB015231.1 product: "stretch-inhibitable nonselective channel (SIC)"; Rattus norvegicus mRNA for stretch-inhibitable nonselective channel (SIC), complete cds., N = 2, Score = 2090, P = 2e-219

>TREMBL:AF029310.1 product: "vanilloid receptor subtype 1"; Rattus norvegicus Vanilloid receptor subtype 1 mRNA, complete cds.  
Length = 838

HSPs:

Score = 3760 (564.1 bits), Expect = 0.0e+00, P = 0.0e+00  
Identities = 721/839 (85%), Positives = 773/839 (92%)

```
Query: 1 MKKWSSTDLGAAADPLQKDTCPDPLGDPNSRPPPAKPQLSTAKSRTLFGKGDSEEAFF 60
M++ +S D + P Q+++C DP D DPN +PPP KP + T +SRTLFGKGDSEEA P
Sbjct: 1 MEQRASLDSEESPPQENSCLDPPDRPNCKPPPVKPHITTSRTLFGKGDSEEAFF 60

Query: 61 VDCPHEEGELDSCPTITVSPVITIQRPDGGPTGARLLSQDSVAASSTKRLRYDRRSIFE 120
+DCP+EEG L SCP ITVS V+TIQRPDGGP R SQDSV+A EK RLYDRRSIF+
Sbjct: 61 LDCPYEEGGLASCPITVSSVLTIQRPDGGPASVRPSSQDSVSAG-EKPPRLYDRRSIFD 119

Query: 121 AVAQNNCOQLESLLFLQKSKKHLTDNEFKDPETGKTKLLKAMLNHOGQNTTIPILLEI 180
AVAQ+NCQ+LESLL FLQ+SKK LTO+EFKDPETGKTKLLKAMLNH+GQN TI LLL+
Sbjct: 120 AVAQSNCOQLESLLFLQKSKKRLTDSEFKDPETGKTKLLKAMLNHNGQNTIALLDV 179

Query: 181 ARQTDLSKELVNASYTDSYKGTALHIAIERNNMALVTLVENGADVQAAANGDFFKKT 240
AR+TDSLK+ VNASYTDSYKGTALHIAIERNNM LVTLLVENGADVQAAA+GDFFKKT
Sbjct: 180 ARKTDLSKQFVNASYTDSYKGTALHIAIERNNMTLVTLLVENGADVQAAANGDFFKKT 239

Query: 241 KGRPGFYFGEPLSLAACTNQLGIVKFLQNSWQTADISARDSVGNVTLHALVEADNTA 300
KGRPGFYFGEPLSLAACTNQL IVKFLQNSWQ ADISARDSVGNVTLHALVEADNT
Sbjct: 240 KGRPGFYFGEPLSLAACTNQLAIVKFLQNSWQADISARDSVGNVTLHALVEADNTV 299

Query: 301 DNTKFTVSMYNEILILGAKLHPTLKLEELTNKKMGTPPLAACTGKIGVLAYILQREIQE 360
DNTKFTVSMYNEILILGAKLHPTLKLEE+TN+KG+TPLALAA +GKIGVLAYILQREI E
Sbjct: 300 DNTKFTVSMYNEILILGAKLHPTLKLEELTNRKGLTPLAALSSGKIGVLAYILQREIHE 359

Query: 361 PECHRLSRKFTWAYGPVHSSLYDLSCIDTCEKNSVLEVIAYSSSETPNRHDMLLVEPLN 420
PECHRLSRKFTWAYGPVHSSLYDLSCIDTCEKNSVLEVIAYSSSETPNRHDMLLVEPLN
Sbjct: 360 PECHRLSRKFTWAYGPVHSSLYDLSCIDTCEKNSVLEVIAYSSSETPNRHDMLLVEPLN 419

Query: 421 RLLQDKWDRFVKRIFYFNFLVYCLYMIIFTMAAYYRPVGLPPFFKMEK-IGDYFRVTGEI 479
RLLQDKWDRFVKRIFYFNFLVYCLYMIIFTMAAYYRPVGLPPFFKMEK+ +GDYFRVTGEI
Sbjct: 420 RLLQDKWDRFVKRIFYFNFLVYCLYMIIFTMAAYYRPVGLPPFFKMEK+ +GDYFRVTGEI 479

Query: 480 LSVLGGVYFFFRGIQYFLQRRPSMKTLFVDSYSEMLFFLQSLFMLATVVLVYFSLKEYVA 539
LSV GGVYFFFRGIQYFLQRRPS+K+LFVDSYSE+LFF+OSLFML +VVLYFS KEYVA
Sbjct: 480 LSVSGGVYFFFRGIQYFLQRRPSLSLFLVDSYSEILFFVQSLFMLVSVVLYFSQRKEYVA 539

Query: 540 SMVFSALGWTNMLYYTRGFQMGYIYAVMIEKMLRDLRCRNFVYIVFLFGFSTAVVTLI 599
SMVFSLA+GWTNMLYYTRGFQMGYIYAVMIEKMLRDLRCRNFVY+VFLFGFSTAVVTLI
Sbjct: 540 SMVFSLANGWTNMLYYTRGFQMGYIYAVMIEKMLRDLRCRNFVYVFLFGFSTAVVTLI 599

Query: 600 EDGKNSLPSSESTSHRWGPACRPDSSYNSLYSTCIELFKFTIGMGDLEFTEYDFKAV 659
EDGKN+SLP EST H+ RG AC+P +SYNSLYSTCIELFKFTIGMGDLEFTEYDFKAV
Sbjct: 600 EDGKNSLPMESTPHKCRGSACKP-GNSYNSLYSTCIELFKFTIGMGDLEFTEYDFKAV 658

Query: 660 FIILLAYVILTYILLNNMLIALMGETVNKIAQESKNWKLQRAITILDTEKSLKCMRK 719
FIILLAYVILTYILLNNMLIALMGETVNKIAQESKNWKLQRAITILDTEKSLKCMRK
Sbjct: 659 FIILLAYVILTYILLNNMLIALMGETVNKIAQESKNWKLQRAITILDTEKSLKCMRK 718

Query: 720 AFRSGKLLQVGYPDGDYRWCFRVDEVNWTNTNNGIINEDPGNCEGVKRTLSFSLR 779
AFRSGKLLQVG+TPDGDYRWCFRVDEVNWTNTNNGIINEDPGNCEGVKRTLSFSLR
Sbjct: 719 AFRSGKLLQVGTPDGDYRWCFRVDEVNWTNTNNGIINEDPGNCEGVKRTLSFSLR 778

Query: 780 SSRVSGRWKFNALVPLLRASARDRQSAQPEEVYLRQFSGSLKPEDAIEVFKSPAASGEK 839
S RVSGR+WKNFALVPLLR+AS RDR + Q EEV L+ ++GSLKPEDAIEVFK GEK
Sbjct: 779 SGRVSGRWKFNALVPLLRDASTDRHATQQEEVQLKHYTGSKPEDAIEVFKSDMVPGER 838
```

Pedant information for DKF2phtes3\_20k2, frame 2

## Report for DKF2phtes3\_20k2.2

[LENGTH] 839  
[MW] 94950.75  
[PI] 6.90  
[HOMOL.] TREMBL:AF029310.1 product: "vanilloid receptor subtype 1"; Rattus norvegicus  
vanilloid receptor subtype 1 mRNA, complete cds. 0.0  
[FUNCAT] 99 unclassified proteins [S. cerevisiae, YIL112w] 4e-05  
[PIRKW] alternative splicing 3e-06  
[PIRKW] peripheral membrane protein 3e-06  
[SUPFAM] ankyrin repeat homology 3e-06  
[SUPFAM] unassigned ankyrin repeat proteins 3e-06  
[PFAM] Ank repeat  
[KW] TRANSMEMBRANE 4

SEQ MKKWSSTDLGAADPLQKDTCPDPLDGPNSRPPPAKPQLSTAKSRTRLFGKGDSEEAFF  
PRD ccc  
MEM .....

SEQ VDCPHEEGELDSCPTITVSPVITIQRPDGPPTGARLLSQDSVAASEKTLRLYDRRSIFE  
PRD ccc  
MEM .....

SEQ AVAQNQCQDLESLLLFLQSKKHLTONEFKDPETGKTCLLKAMLNLDGQNTTIPLLLEI  
PRD hhhhechhhhhhhhhhhhhcc  
MEM .....

SEQ ARQDLSKELVNASYTDVYKGTALHIAIERRMALVTLLVENGADVQAAAHGDFFKKT  
PRD hhcc  
MEM .....

SEQ KGRPGFYFGELPLSLAACTNQLGIVKFLQNSWQTADISARDSVGNLVLHALVEADNTA  
PRD ccc  
MEM .....

SEQ DNTKFVTSMYNEILILGAKLHPTLKLLELTNKKGMTPLAAGTGKIGVLAYILQREIQE  
PRD chhhhhhhhhhhhhhhcc  
MEM .....

SEQ PECRHLSRKFTWYAGPVHSSLYDLSCIDTCEKNSVLEVIAYSSSETPNRHDMLEVEPLN  
PRD ccc  
MEM .....

SEQ RLLQDKWDRFVKRI FYFNFLVYCLYMIIFTMAAYRVPDGLPPFKMEKIGDYFRVTGEIL  
PRD hhh  
MEM .....

SEQ SVLOGVYFFFRGQYFLQRRPSMKTLEVDYSSEMLFFLOSFLMLATVVLVYFSLKEYVAS  
PRD ccc  
MEM .....

SEQ MVFSLALGWTNMLYYTRGFQMGIVAVMIEKMLRLDLRFMFVYIVFLFGFSTAVVTLIE  
PRD hhh  
MEM .....

SEQ DGKNDLSESTSHRWGFPACRPDSSYNLSYSTCLELKFKTIGMGDLFTENYDFKAVF  
PRD ccc  
MEM .....

SEQ ITLLAYVILTYILLNMLIALMGETVKNIAQESKNIWKLQRAITILDEKSFLLKCMRKA  
PRD hhh  
MEM .....

SEQ FRSGKLLQVGYTPDGKDDYRWCFRVDEVNMTTWNINVGIIINEDPGNCEGVKRTLSFSLRS  
PRD hhcc  
MEM .....

SEQ SRVSGRWKKNFALVPLREASARDQSAQPEEYILRQFSGSLKPEDAEVFKSPAASGEK  
PRD ccc  
MEM .....

(No Prosite data available for DKF2phtes3\_20k2.2)

Pfam for DKF2phtes3\_20k2.2



WO 01/12659

PCT/IB00/01496

HMM_NAME	Ank repeat
HMM	*GyTPLHIAARYNNVEMVc1LLQHGADIN*
	G+T+LHIA +++N+ +V LL+++GAD+
Query	202 GQTALHIAIERNNMALVTLLVENGADVQ 229

WO 01/12659

PCT/IB00/01496

DKF2phtes3\_2013  
-----

group: transmembrane protein

DKF2phtes3\_2013 encodes a novel 595 amino acid protein with partial similarity to the IL-17 receptor.

The novel protein contains one transmembrane region.  
No informative BLAST results; No predictive prosite, pfam or SCOP motif.

The new protein can find application in studying the expression profile of testis-specific genes and as a new marker for testicular cells.

similarity to IL-17 receptor

Sequenced by MediGenomix

Locus: unknown

Insert length: 2406 bp  
Poly A stretch at pos. 2345, no polyadenylation signal found

```
1 GCCTCAGGTG TTCTCGGTT GTTTGTGAGT GGAGAGCAGG GAGTGGGGCC
51 AGCCAGCAGA AACAGTGGGC TGTACAACAT CACCTTCAAA TATGACAATT
101 GTACCACCTA CTTGAATCCA GTGGGGAAGC ATGTGATTGC TGACGCCACG
151 AATATCACCA TCAGCCAGTA TGCTTGCCAT GACCAAGTGG CAGTCACCAT
201 TCTTTGGTCC CCAGGGGCCC TCGGCATCGA ATTCTGAA GATTTCGGG
251 TAATACTGGA GGAGCTGAAG TCGGAGGAA GACAGTCCCA ACAACTGATT
301 CTAAGGATC CGAAGCAGCT CAACAGTAGC TTCAAAAGAA CTGGAATGGA
351 ATCTCAACCT TTCTTCAATA TGAAATTTGA AACGGATTAT TTGCTAAAGG
401 TTGTCCCTTT TCCTTCCATT AAAACGAAA GCAATTACCA CCCTTTCTTC
451 TTTAGAACCC GAGCCTGTGA CCTGTTGTTA CAGCCGACA ATCTAGCTTG
501 TAAACCCCTC TGGAAGCCTC GGAACCTGAA CATCAGCAG CATGGCTCGG
551 ACATGCAGGT GTCTCTCGAC CACGACCCGC ACACTTCGG CTTCGGTTTC
601 TTCTATCTTC ACTACAAGCT CAAGCACGAA GGACCTTTCA AGCGAAAGAC
651 CTGTAAGCAG GAGCAAACTA CAGAGATGAC CAGCTGCCTC CTCAAAATG
701 TTTCTCCAGG GATTATATA ATTGAGCTGG TGGATGACAC TACACACACA
751 AGAAAGTGA TCATTATGCT CTTAAAGCCA GTGCACTCCC CTTGGCGCGG
801 GCCCATCAGA GCCGTGGCCA TCACAGTGCC ACTGTAGTCT ATATCGGCAT
851 TCGCGACGCT CTTCACTGTG ATGTGCCGCA AGAAGCAACA AGAAATATA
901 TATTACATT TAGATGAAGA GAGCTCTGAG TCTTCCACAT AACTGCAGC
951 ACTCCAAGA GAGAGGCTCC GGCCGCGGCC GAAGTCTTT CTCTGCTATT
1001 CCAGTAAGA TGCCAGAAAT CACATGAATG TCGTCCAGTG TTTCGCTTAC
1051 TTCTCCAGG ACTTCTGTGG CTGTGAGGTG GCTCTGGACC TGTGGGAAGA
1101 CTTGAGCCTC TGTAGAGAAG GGCAGAGAGA ATGGGTGATC CAGAAGATCC
1151 ACGAGTCCCA GTTCATCATT GTGGTTTGT CCAAGGTAT GAAGTACTTT
1201 GTGGACAAGA AGAATACAAA ACACAAAGGA GGTGCGCAG CTTGGGGAA
1251 AGCAGAGCTC TTCTGTGTGG CGGTGTGAGC CATTGCCGAA AGCTCCGCC
1301 AGGCCAAGCA GAGTTGCTCC GCGGCCCTCA GCAAGTTTAT CGCGTCTAC
1351 TTTGATTATT CTGCGAGGGG AGACGTCCCC GGTATCTTAG ACCTGAGTAC
1401 CAAGTACAGA CTCATGGACA ATCTTCTCTA GCTCTGTTCC CACCTGCATC
1451 CCCGAGACCA CGGCTCCAGG GAGCCGGGGC AGCACACGCG ACAGGGCAGC
1501 AGAAGGAAC ACTTCCGGAG CAAGTCAGGC CGGTCCCTAT ACGTCGCCAT
1551 TTGCAACATG CACCAAGTTA TTGACGAGGA GCCCGACTGG TTGAAAGAGC
1601 AGTTCGTTCC CTTCCATCCT CCTCCACTGC GCTACCGGGA GCCAGTCTTG
1651 GAGAAATTTG ATTCCGGCTT GGTTTTAAAT GATGTCATGT CCAACCGGG
1701 GCTGAGAGT GACTCTGCTC TAAAGTAGA GCGGCTGTT CTTGGGGCNA
1751 CCGGACAGC CGACTCCAG CACGAGAGTC AGCATGGGGG CTTGGACCAA
1801 GACGGGAGG CCGGCTCGC CTTGACGGT AGCGCCGCC CTTGAAACCCCT
1851 GCTGCACAG GTGAAAGCCG GCAGCCCTC GGACATGCG CGGACTCAG
1901 GCATCTATGA CTCGTCTGTG CCTCATCCG AGCTGTCTCT GCCACTGATG
1951 GAAGGACTCT CGACGGACCA GACAGAAACG TCTTCCCTGA CGGAGAGCGT
2001 GTCTCTCTCT TCAGGCTGG GTGAGGAGGA ACCTCCTGCC CTTCTTCCA
2051 AGCTCCTCTC TTCTGGGTCA TGCAAGCAG ATCTTGGTTG CCGCAGCTAC
2101 ACTGATGAAC TCCACGGGTG CGCCCCCTTG TAAACAAACG AAAGAGTCTA
2151 AGCATTGCCA CTTTAGCTGC TGCCCTCCCTC TGATTCCCA GCTCATCTCC
2201 CTGGTTGCAT GGGCCACTTG GAGCTGAGGT CTATACAG GATATTGGA
2251 GTGAATGCT GGGCAGTACT TGTCTCCCT TGCCCAACC CTTTACCGGA
2301 TATCTTGACA AACTCTCAA TTTCTAAAA TGATATGAG CTCTGAAAAA
2351 AAAAAA AAAA AAAA AAAA AAAA
2401 AAAAAA
```

BLAST Results  
-----

No BLAST result

## Medline entries

No Medline entry

## Peptide information for frame 1

ORF from 346 bp to 2130 bp; peptide length: 595  
Category: similarity to known protein  
Classification: unclassified

```
1 MESQPLNMK FETDYFVKV PPSIKNESN YHPFFFRTRA COLLQPDNL
51 ACKPFWKPRN LNISQHGSDM QVSFDHAPHN FGFRFFYLHY KKHGEPFKR
101 KTKQEQTTE MTSCLLQNVN PGDYIIELVN DNTTRKVMH YALKPVHSPW
151 AGPIRAVAIT VPLVVISAFN TLFVVMCRKK QQENIYSHLD EESSESSTYT
201 AALPRERLRP RPKVFLCYSS KDGQNMHNVV QCFATFLQDF CGCEVALDLN
251 EDFSLCRGQJ RFWIQRKHE SQTIVVCSK QMKYFVDKKN YKHKGGGRGS
301 GKGEFLVAV SAIAEKLRQA KQSSAALSX FIAVYFDYSC EGDVPGILD
351 STKYLMDNL POLCSHLHSR DHGLQEPQGH TROGRRNYF RSKSGRSLYV
401 AICNMHQFID EEPDWFQKF VPFNPPPLRY REPVLEKFDG GLVLNDVMCK
451 PGFESDFCLK VEAALVLTGK PADSQHSQH GGLDQGEAR PALDGSAAQ
501 PLLHTVKAGS PSDMPSRSGI YDSSVPSSEL SLPLMEGLST DQETSSLTE
551 SVSSSSGLGE EEPALPSKL LSSGCKADL GCRSYTDELH AVAPL
```

## BLASTP hits

No BLASTP hits available

## Alert BLASTP hits for DKFZphtes3\_2013, frame 1

TREMBL:U58917.1 product: "IL-17 receptor"; Homo sapiens IL-17 receptor  
mRNA, complete cds., N = 1, Score = 215, P = 4.7e-14

TREMBL:MM31993.1 product: "interleukin 17 receptor"; Mus musculus  
interleukin 17 receptor mRNA, complete cds., N = 2, Score = 152, P =  
1.1e-13

>TREMBL:U58917.1 product: "IL-17 receptor"; Homo sapiens IL-17 receptor  
mRNA, complete cds.  
Length = 866

## HSPs:

Score = 215 (32.3 bits), Expect = 4.7e-14, P = 4.7e-14  
Identities = 85/284 (29%), Positives = 131/284 (46%)

```
Query: 213 KVFLEYSSKDGQNMHNVVQCFAYFLQDFCGCEVALDLMEDFSLCRGQREWV-IQK---I 268
      KV++ YS+ D +++VV FA FL CG EVALDL E+ ++ G WV QK +
Sbjct: 379 KVMIIYSA-DHPLYVDVVLKFAQFLLTACGTEVALDLLEEQAISEAGVMTWVGQRKQEMV 437

Query: 269 HESQFIIVVCSKGMKY----FVDKKNYXXXXXXXXXXELFLVAVSAIAEXXXXXXXXX 324
      + IIV+CS+G + + + +LF A++ 1
Sbjct: 438 ESNKIIIVLCRGRTRAKWQALLGRGAPVRLRCDHGKPVGDLFTRAMNHILPDFKRPACFG 497

Query: 325 XXXXXXFIAYVF-DYSCGQDVPGLDLSTKYRLMDNLPOLCSHLHSRDHGLQEPGQHTRO 383
      ++ YF + SC+GDVP + + +Y LMD ++ + +D + +PG+ R
Sbjct: 498 T-----YVVCYFSEVSCDGDVPDLFGAAPRYPLMDRFEEV--YFRIQDLEMFPQGRMHRV 550

Query: 384 G--SRNRYFRSKSGRSLYVAICNMHQFIDEEPWFQKQFV----PFNPPPLR---YREP 434
      G S NY RS GR L A+ + PDWFE + + P L + EP+
Sbjct: 551 GELSGDNYLRSPGGRLRAALDRFRDQVRCQDWFECENLYSADDQDAFSLDEEVFEPL 610

Query: 435 LEKFDSGLVLNDVMCKPGPESDFCLKVEAAVLGATGPADSQHSQHGLDQDGEARP 491
      L +G+V + + P S CL++ V G GA H L G+ P
Sbjct: 611 LPP-GTIVKRAPLVRE-PGSACLAIDPLV-GEAGAAVAKLEPH--LQPRGQAP 662
```

## Pedant information for DKFZphtes3\_2013, frame 1

Report for DKFZphtes3\_2013.1

```

2551 ACGCAAGATT GCCAAATTCG AGGAGAAGCA CTTGTCGAGT TTAAGTGCCA
2601 TTCGAGAGGA GTTGGAACTG CCCAACATTG AGAAGATGAT CCTAGAATGC
2651 AGTGCTGACA TCAGTGAGTT GTTCGATGCG CTCATGACGC TGGAGATGCA
2701 GCTGGTGGAG CAGCTGGAGG TAAGGCTGGG CCCTGGGCAC AAGTGCCAGA
2751 ATCTGGCGAT GCAGCTGCAC ATCCATAGGT GAACTGTAGC CTTCATGGGC
2801 ACGCCTCTGC TGGAAACGTC CAGCAGACT CAGCGTGGCA GGCTGTAGCT
2851 TTCTTGCTCA TCAGTCCTGT TTGCTTTTAT TACATTTTAA TCATTTACAT
2901 TGGAGTGAT TCTTGTGGAA AATGAGAGGT GAGCTCATTC TTCTGAAATG
2951 GTCCCCCTAT CCTGGAAATC AGTGGGGAGA GGTTTTTGAT TAGACCCCTG
3001 GAGCTATCCG GGTACTCTAA AGGCAAAGCG CACCCCACT TGGGGACCAA
3051 ACAAGAGACC CTCCGCATTG CAGCCTGCAG TTGCCGCTTC TCAGGTGACG
3101 TGAGGAGGCT GCAACTCAGC ACTAAGTAGT GAAAATGAAA AGCGCCGCTG
3151 TCTGAAATTC ATTAGCAGCC AGAGTATGTG TTACAAGGCA GCGGAGGCTG
3201 GGAATCTGAA GTGGTGTGAT GAATTGAACC TCATCGGATG CTGCTGTGGC
3251 TGGGCCAAGT GATAGCACT AATCAATTCC TCACACGTCA AGTGACACCT
3301 CAGACATGGG ATAGATTTCC CCATCACATC ACAGGGCAGG TGCTCCCTCC
3351 CTGCTGGAGA GCACAGGCAC TGCAGAAGCA GCGCACAGTG CCAGGGGCGA
3401 GTGAGGCAGC AGCTCCAGC CTTTTCAGGC ACGGAGATTG CCTTTCAACA
3451 TCCAAACATT TCCGAGAACC CATGTGCCAT CCTACTTGTA TTAAGTGGT
3501 CCAGAAAGCC ACAAGCGCAA TCATGCTTTT CAATGACCCT ATTTTATTTC
3551 ACGAGAACAG CACATACATG TGTTTGAAAA TTATGTGAGG TGCTCATCTT
3601 GCAGACAGTA CTCACATTCC TATAGATTCC ACCCTGCCCC ACCTTGCAGC
3651 CCCTGGAGTC TATAGCAGAT GGGAGTGGGG CACTCCGAGA GTGGCAGGCC
3701 TGGAGATCAC ATCTTCCATT GTTCCTTCAA TCAACACTAA CTCCCATTTG
3751 GGCTTAGGTG GCCTTGCTAA GCACCACAAA ACAGCAACTA ACTGAAAGAG
3801 ATCTGGAGTG CAGCCCGCT CACTACTGAGG GCCTCCTCTC TGTGAGGCAC
3851 CTTGCAAAGC ATTTTGTGTG AAGTGACTCA TTTAACCTCA CCACAACGCC
3901 ACAACGCAGG GATTATGCAG GTAACCTATT TCCCAGATGA GGAAGATAAG
3951 GCCCAGAGAG GTGAAATGCC TTTCACAGAG TTACACAGAG TGCTGGAGCT
4001 GGAATACTG ACCCAGGCAG TCTAGCTCTT AACAGCTCAC TCCACTGTTT
4051 CCCTGGAGGT GATGCACAGA TGCTACTGGG AAACCCAAAG GAGAGGGGGT
4101 TGGCTGTGTG TGTGTGTGTT GGGCAGGCAG GTAAGGGGAG TAAGACCAAG
4151 ACAAGTGTTC CTGGCAAAGT TCCGGTGACA GCATTAAACA TTCAGATGGT
4201 GAGGGAGTTA ATATGGTTGG AGAACAACAA CTTTAGAGAG AGCAGAGGGG
4251 TCAGTTTACA ACCATCTGCT CAGGAGGGTC AAGATGGGTG GTCTTTATGC
4301 TGAAGGCTG TGATTAGAGG AGCTGGTTGC TAAATTTTGA GGAGTACCTT
4351 TTGCTCTGTG CTGGACATCT AAATATGCAT GTTAACGTG TTCTTTAACA
4401 TTTCCAGGAG ACTATAACA TGTTTGAAAG GAACATTGTT GACATGGTAG
4451 GACTGTTTAT CGAAAATGTC CAAAGCCTAT ATCCTTTCTG TGATGACCTT
4501 CCCCATGGGG AGGTGCTACA GAGCCCTGG GCTTGTCCTG GCCTCTGGAC
4551 AAAAGAATGT TCCACAGGGT CTGAGGAGGT TTCCCGACCC TCAGAACAAT
4601 GATGGCCTGG TTAGAGCTGT GGTTTGGATG CCCAGAGGGA CAACATCCAA
4651 ACTGTTTGCA GTAGGCTCCC AGCATGATTG TTCTCATATG AGTGATGTTT
4701 ACTAGGAAAT GACGCCCCCT GTGTTGCAGG CAAGCACACT CTGGGGTTGA
4751 GGCAACCCCC ACGTGGGAAG CACTATAAAG AGTACATCAG GTGAAATGTT
4801 AGGGTGAGGA GCCAACATCG GAGCATGGCC AACCTTCTT CCACCCGAAC
4851 TCAGGGCACT CCACATGGGG CAAACTGCTG TGCTCCAGCT AGCAGCAGCC
4901 CTGTGGTCCT GCCCTCCTGG GGCTCACAGT CCCTCAGGGA GACAAGTTGT
4951 AGAGGCAACA AGTGGTGCCA AATGCACAGG GTGAGAAGCA GTTAACCCAG
5001 AGGCCAGGAG CCTCCATGCA GGAGGGAGAG AAGAGTGTGA TGGCAGGGGC
5051 CGAGGGTCCG TCCGAGGTGT GGGGCAGGGG CAGGGAGTCG AGGAAGGCCC
5101 AGGGTTCGGA GCTTGTGAGT GGACGGTGCT GCCAGCCAGA ATTTCCGAGC
5151 TCGCCTTGGG CCTTAAAGT CTGCTCCTCC CCGTCTGAGA GCATCAGGGA
5201 CGCGCCGGG CTGCTCCTCC CGGCCTTTG CTTAACTCGG GGCTGCACGA
5251 TGGCTCAGTG CCGGGACCTG GAGAATCACC ACCACGAGAA GCTCCTGGAG
5301 ATCTCTATCA GCACCTGGA GAAGATTGTC GAGGGCGACC TGGACGAGGA
5351 CCTGCTTAAC GACCTGCGCG CGCTTTTGT CGATAAAGAT ACGATTGTTA
5401 ATGCTGTCGG GGCATCGCAC GACATCCACC TCCTGAAGAT TGACAATCGA
5451 GAAGATGAGC TGGTGACCAG AATCAACTCT TGGTGACAC GTTTAATAGA
5501 CAGGATTAC AAGGATGAGA TCATGAGGAA CCGCAAGCGC GTGAAGGAGA
5551 TCAATCAGTA CATCGACCAC ATGCAGAGCG AACTGGACAA CCTGGAATGT
5601 GGCACATCC TAGACTAGAT GAATGTCAGC CACAGGAGCT TCTTCAAAAC
5651 ATAGCACCA GAGAGGGAAG TGACACGCGC TCACCCGCAC
5701 CTCTAGAGAG TTGCTGGGCA TCTCTCAACC GCGATCCCCA ACACCATCTT
5751 TCCCCACCC CTGGAAAAAC TTCCAAAAGT AGAGAAAATA AAGGACTCAT
5801 TTCACAAAAA AAAAAAAAAA AA

```

# BLAST Results -----

Entry HSI292248 from database EMBL:  
human STS SHGC-53917.  
Score = 874, P = 3.3e-33, identities = 180/185

# Medline entries -----

No Medline entry

## Peptide information for frame 1

ORF from 202 bp to 876 bp; peptide length: 225  
Category: similarity to known protein

```

1 MNQPCNSMEP RVMDDDLKLL AVGDQGPQEE AGQLAKQEGI LFKDVLSLQL
51 DFRNLRIDN LWQFENLRKL QLDNNIEKI EGLENLAHLV WLDLSFNNIE
101 TIEGLDTLVN LEDLSLFNNR ISKIDSLDAL VKLQVLSLGN NRIDNMMNII
151 YLRRFKCLRT LSLSRNPIS EADYKMFICA YLPDLMYLDY RRIDDHTASV
201 SLSVSQPCET DSSSPQVSWK RGIEE

```

## BLASTP hits

Entry S68209 from database PIR:  
sds22 protein homolog - human >TREMBL:HSSDS22MR.1 gene: "sds22";  
product: "yeast sds22 homolog"; H.sapiens sds22-like mRNA  
Score = 234, P = 1.2e-19, identities = 61/143, positives = 93/143

Entry A38439 from database PIR:  
suppressor protein sds22(+) - fission yeast (Schizosaccharomyces pombe)  
>TREMBL:SPSDS22.1 gene: "sds22+"; S.pombe sds22+ gene, complete cds.  
Score = 208, P = 5.6e-17, identities = 52/127, positives = 71/127

Entry S43988 from database PIR:  
protein suppressor sds22 - fission yeast (Schizosaccharomyces pombe)  
>SWISSPROT:SD22 SCHPO PROTEIN PHOSPHATASES PP1 REGULATORY SUBUNIT  
SDS22. >TREMBL:SPAC4A8.12 gene: "sds22"; product: "phosphatases pp1  
regulatory subunit"; S.pombe chromosome I cosmid c4A8.  
Score = 208, P = 8.5e-17, identities = 52/127, positives = 71/127

Entry CEK10D2.5 from database TREMBL:  
gene: "K10D2.1"; Caenorhabditis elegans cosmid K10D2.  
Score = 214, P = 3.6e-16, identities = 50/125, positives = 75/125

## Alert BLASTP hits for DKFZphut1\_20ml1, frame 1

No Alert BLASTP hits found

## Pedant information for DKFZphut1\_20ml1, frame 1

## Report for DKFZphut1\_20ml1.1

```

[LENGTH]      225
[MW]           25955.87
[pI]           4.63
[HOMOL]        PIR:S68209 sds22 protein homolog - human 1e-18
[FUNCAT]        03.22 cell cycle control and mitosis [S. cerevisiae, YKL193c] 2e-11
[FUNCAT]        30.10 nuclear organization [S. cerevisiae, YKL193c] 2e-11
[FUNCAT]        06.07 protein modification (glycosylation, acylation, myristylation,
palmitoylation, farnesylation and processing) [S. cerevisiae, YKL193c] 2e-11
[FUNCAT]        30.05 organization of centrosome [S. cerevisiae, YOR373w] 2e-06
[FUNCAT]        01.03.10 metabolism of cyclic and unusual nucleotides [S. cerevisiae,
YJL005w] 3e-05
[FUNCAT]        03.10 sporulation and germination [S. cerevisiae, YJL005w] 3e-05
[FUNCAT]        30.02 organization of plasma membrane [S. cerevisiae, YJL005w] 3e-05
[FUNCAT]        10.04.03 second messenger formation [S. cerevisiae, YJL005w] 3e-05
[FUNCAT]        04.07 rna transport [S. cerevisiae, YPL169c] 9e-04
[FUNCAT]        04.05.01.04 transcriptional control [S. cerevisiae, YCR065w] 9e-04
[EC]            4.6.1.1 Adenylate cyclase 2e-06
[PIRKW]         nucleus 5e-16
[PIRKW]         duplication 2e-06
[PIRKW]         tandem repeat 2e-06
[PIRKW]         cAMP biosynthesis 2e-06
[PIRKW]         glycoprotein 2e-06
[PIRKW]         phosphorus-oxygen lyase 2e-06
[SUPFAM]        leucine-rich alpha-2-glycoprotein repeat homology 5e-16
[SUPFAM]        fibromodulin 3e-07
[SUPFAM]        yeast adenylate cyclase catalytic domain homology 2e-06
[SUPFAM]        yeast adenylate cyclase 2e-06
[PROSITE]       CK2_PHOSPHO_SITE 2
[PROSITE]       PKC_PHOSPHO_SITE 1

```

[KW] All\_Alpha

```

SEQ  MNQPCNSMEPRVMDDDLKLAAGDQGPQEEAGQLAKQEGILFKDVLSLQLDFRNILRIDN
PRD  cccccccccccccchhhhhccccchhhhhhhhhhhchhhhhhhhhcccccccccccc

SEQ  LWQFENLRKLQLDNNIIEKIEGLENLAHLVWLDLSFNNIETIEGLDTLVNLEDLSLFNNR
PRD  hhhhhhhhhhhccccccccccccchhhhhhhccccccccccccchhhhhhhhhcccccc

SEQ  ISKIDSLDALVKLVLSLGNRRIDNMMNIIYLRREFKCLRTLSSRNPISEAEDYKMFICA
PRD  cccchhhhhhhhhhhccccccccccccccccchhhhhhhhhccccccccchhhhhhhhh

SEQ  YLPDLMYLDYRRIDDHTASVLSVSPQPCETDSSSPQVSWKRGIEE
PRD  hhccccccccccccchhhhhhhcccccccccccccccccccccc

```

## Prosites for DKFZphut1\_20m11.1

PS00005	218->221	PKC_PHOSPHO_SITE	PDOC00005
PS00006	122->126	CK2_PHOSPHO_SITE	PDOC00006
PS00006	169->173	CK2_PHOSPHO_SITE	PDOC00006

(No Pfam data available for DKFZphut1\_20m11.1)

DKFZphute1\_20m24

group: metabolism

DKFZphute1\_20m24 encodes a novel 611 amino acid protein with similarity to a hypothetical *C.elegans* protein and to yeast Alg9 protein.

This protein is a putative mannosyl transferase that is involved in the assembly of the core oligosaccharide Glc3Man9GlcNAc2.

The new protein can find application in modulation of glycosylation of proteins and as a new enzyme for biotechnologic production processes.

strong similarity to *S.cerevisiae* Alg9p

complete cDNA, complete cds, potential start at Bp 23, few EST hits  
Alg9 is involved in the assembly of the core oligosaccharide  
Glc3Man9GlcNAc2

HSAC381 corresponding genomic DNA (2 exons)

HSB8954 corresponding genomic DNA (1 exon)

Sequenced by AGOWA

Locus: /map="11"

Insert length: 1986 bp

Poly A stretch at pos. 1966, polyadenylation signal at pos. 1949

```

1 TTCTTTTTC CCCAGGCTTG CCATGGCTAG TCGAGGGGCT CGGCAGCGCC
51 TGAAGGGCAG CGGGGCCAGC AGTGGGGATA CGGCCCGGCG TGCGGACAAG
101 CTGCGGGAGC TGCTGGGCAG CCGAGAGGCG GCGGGCGCGG AGCACC GGAC
151 CGAGTTATCT GGGAAACAAAG CAGGACAAGT CTGGGCACCT GAAGGATCTA
201 CTGCTTTCAA GTGTCTGCTT TCAGCAAGGT TATGTGCTGC TCTCCTGAGC
251 AACATCTCTG ACTGTGATGA AACATTCAAC TACTGGGAGC CAACACACTA
301 CCTCATCTAT GGGGAAGGGT TTCAGACTTG GGAATATTCC CCAGCATATG
351 CCATTCGCTC CTATGCTTAC CTGTGTGCTT ATGCCTGGCC AGCTGCATTT
401 CATGCAAGAA TTCTACAAAC TAATAAGATT CTTGTGTTTT ACTTTTTGCG
451 ATGTCTTCTG GCTTTTGTGA GCTGTATTTG TGAACCTTAC TTTTACAAGG
501 CTGTGTGCAA GAAGTTTGGG TTGCACGTGA GTCGAATGAT GCTAGCCTTC
551 TTGGTTCTCA GCACTGGCAT GTTTTGCTCA TCATCAGCAT TCCTTCCTAG
601 TAGCTTCTGT ATGTACACTA CGTTGATAGC CATGACTGGA TGGTATATGG
651 ACAAGACTTC CATGTGCTGT CTGGGAGTAG CAGCTGGGGC TATCTTAGGC
701 TGGCCATTCA GTGCAGCTCT TGGTTTACCC ATTGCCTTGG ATTTGCTGGT
751 CATGAAACAC AGGTGGAAGA GTTTCTTTCA TTGGTCGCTG ATGGCCCTCA
801 TACTATTCTT GGTGCCTGTG GTGGTCATTG ACAGCTACTA TTATGGGAAG
851 TTGGTGATTG CACCACTCAA CATTGTTTTG TATAATGTCT TTA CTCTCA
901 TGGACCTGAT CTTTATGGTA CAGAACCCTG GTATTCTTAT TTAATTAATG
951 GATTCTTGAA TTTCAATGTA GCCTTTGCTT TGGCTCTCCT AGTCCTACCA
1001 CTGACTTCTC TTATGGAATA CCTGCTGCAG AGATTTCATG TTCAGAAATT
1051 AGGCCACCCG TATTGGCTTA CCTTGGCTCC AATGTATATT TGGTTTATAA
1101 TTTTCTTCAT CCAGCCTCAC AAAGAGGAGA GATTTCTTTT CCCTGTGTAT
1151 CCACTTATAT GTCTCTGTGG CGCTGTGGCT CTCTCTGCAC TTCAGAAATG
1201 TTACCACTTT GTGTTTCAAC GATATCGCCT GGAGCACTAT ACTGTGACAT
1251 CGAATTGGCT GGCATTAGGA ACTGTCTTCC TGTTTGGGCT CTTGTCATTT
1301 TCTCGCTCTG TGGCACTGTT CAGAGGATAT CACGGGCCCC TTGATTTGTA
1351 TCCAGAATTT TACCGAATTG CTACAGACCC AACCATCCAC ACTGTCCCAG
1401 AAGGCAGACC TGTGAATGTC TGTGTGGGAA AAGAGTGGTA TCGATTTCCT
1451 AGCAGCTTCC TTCTTCCTGA CAATTGGCAG CTTCAAGTTC TTCCATCAGA
1501 GTTCAGAGGT CAGTTACCAA AACCTTTTGC AGAAGGACCT CTGGCCACCC
1551 GGATTGTTCC TACTGACATG AATGACCAGA ATCTAGAAGA GCCATCCAGA
1601 TATATTGATA TCAGTAAATG CCATTATTTA GTGGATTGTT ACACCATGAG
1651 AGAAACACCC CGGGAGCCAA AATATTCATC CAATAAAGAA GAATGGATCA
1701 GCTTGGCCTA TAGACCATTC CTTGATGCTT CTAGATCTTC AAAGCTGCTG
1751 CGGGCATTCT ATGTCCCCTT CCTGTCAGAT CAGTATACAG TGTACGTAAA
1801 CTACACCATC CTCAAACCCC GGAAAGCAAA GCAATCAGG AAGAAAAGTG
1851 GAGGTTAGCA ACACACCTGT GGCCCCAAG GACAACCATC TTGTTAACTA
1901 TTGATTCCAG TGACCTGACT CCCTGCAAGT CATGCCTGT AACATTTGTA
1951 ATAAAGGTCT TCTGACATGA AAAAAAAAAA AAAAAA
```

## BLAST Results

Entry HSAC381 from database EMBL:  
Homo sapiens chromosome 11 pac pDJ159o1, complete sequence.  
Length = 42,771

Entry HSB8954 from database EMBL:

cSRL-50A3-u cSRL flow sorted Chromosome 11 specific cosmid Homo sapiens genomic clone cSRL-50A3.  
Length = 601

# Medline entries

96293493:  
Stepwise assembly of the lipid-linked oligosaccharide in the endoplasmic reticulum of *Saccharomyces cerevisiae*: identification of the ALG9 gene encoding a putative mannosyl transferase.

# Peptide information for frame 2

ORF from 23 bp to 1855 bp; peptide length: 611  
Category: strong similarity to known protein

```

1 MASRGARQRL KGSGASSGDT APAADKLREL LGSREAGGAE HRTLSGNKA
51 GQVWAPEGST AFKCLLSARL CAALLSNISD CDETFNYWEP THYLIYGEFG
101 QTWEYSPAYA IRSYAYLLH AWPAAFHARI LQTNKILVEY FLRCLLAFVS
151 CICELYFFYKA VCKKFGHLVS RMMLAFLVLS TGMFCSSSAF LPSSFCMYTT
201 LIAMTGWYMD KTSIAVLGVA AGAILGWPF S AALGLPIAFD LLVMKHRWKS
251 FFHWSLMALI LFLVPVVVID SYYYGKLVIA PLNIVLYNVF TPHGPDLYGT
301 EPWYFYLING FLNFNVAFAL ALLVLPLTSL MEYLLQRFHV QNLGHPYWLT
351 LAPMYIWFII FFIQPHKEER FLFPVYPLIC LCGAVALSAL QKCYHFVFQR
401 YRLEHYTVTS NWLALGTVEL FGLLSFSRSV ALFRGYHGPL DLYPEFYRIA
451 TDPTIHTVPE GRPVNVCVGK EWYRFPSSFL LPDNWQLQFI PSEFRGQLPK
501 PFAEGPLATR IVPTDMNDQN LEEPSRYIDI SKCHYLVLDL TMRETPREP
551 YSSNKEEWS LAYRPFLLDAS RSSKLLRAFY VFFLSDQYTV YVNYTILKPR
601 KAKQIRKKSG G

```

# BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphutel\_20m24, frame 2

SWISSPROT:YTH3\_CAEEL HYPOTHETICAL 75.5 KD PROTEIN C14A4.3 IN CHROMOSOME II., N = 1, Score = 957, P = 2.7e-96

PIR:S63177 mannosyl transferase (EC 2.4.1.-) - yeast (*Saccharomyces cerevisiae*), N = 1, Score = 533, P = 2.3e-51

SWISSPROT:YTH3\_CAEEL HYPOTHETICAL 75.5 KD PROTEIN C14A4.3 IN CHROMOSOME II., N = 1, Score = 957, P = 2.7e-96

PIR:S63177 mannosyl transferase (EC 2.4.1.-) - yeast (*Saccharomyces cerevisiae*), N = 1, Score = 533, P = 2.3e-51

>SWISSPROT:YTH3\_CAEEL HYPOTHETICAL 75.5 KD PROTEIN C14A4.3 IN CHROMOSOME II.

Length = 653

# HSPs:

Score = 957 (143.6 bits), Expect = 2.7e-96, P = 2.7e-96  
Identities = 206/514 (40%), Positives = 296/514 (57%)

```

Query: 48 NKAGQVWAPEGSTAFKCLLSARLCAALLSNISDCDETFNYWEPHTHYLIYGEFGQTWEYSP 107
      N   W   + FK LLS R+ A+ I+DCDE +NYWEP H +YGEFGQTWEYSP
Sbjct: 43 NNPDNDWPFSFGSVFKMLLSIRISGAIWGIINDCDEVYNYWEPLHLFLYGEFGQTWEYSP 102

Query: 108 AYAIRSYAYLLHAWPAAFHARILQTNKILVEYFLRCLLAFVSCICELYFYKAVCKKFG 167
      YAIRSY Y+ LH PA+ A + KI+VF +R + + E Y + A+CKK +
Sbjct: 103 VYAIRSYFYIYLHYIPASLFANLFGDTKIVFTLIRLTIGLFCLLGEYAFDAICKKINI 162

Query: 168 HVSRRMLAFLVLSTGMFCSSSAFLPSSFCMYTTLIAMTGWYMDKTSIAVLGVAAGAILGW 227
      R + F + S+GMF +S+AF+PSSFCM T + + + + + VA ++GW
Sbjct: 163 ATGRFFILFSIFSSGMFLASTAFVPSFSCMAITFYILGAYLNENWTAGIFCVAFSTMVGW 222

Query: 228 PFSAALGLPIAFDLLVMKHRWKSFFHWSLMALILFLVPVVVIDSYYYGKLVIAPLNIVLY 287

```



Sbjct: 223 PFSA LGLPI D+L++K F SL+ + V+ DS+Y+GK V+APLNI LY  
 Query: 288 NVFTPHGPDLYGTEPWYFYLINGFLNFNVAFA LALLVPLTSLMEYLLQRFHVQNLGHPY 347  
 NV + GP LYG EP FY+ N F N+N+ A PL+ + Y + + Q+  
 Sbjct: 283 NVVSGPGPSLYGEEPLSFYIKNLFNNWNIVIFAAPFGFPLS--LAYFTKVWMSQDRNVAL 340  
 Query: 348 WLT LAPMYI-----WFIIFFIQPHKEERFLFPVYPLICLCAVALSALQKCYHVFQR 400  
 + AP+ + W +IF Q HKEERFLFP+YP I A+AL A + ++  
 Sbjct: 341 YQRFAPIIILLAVTTAAWLLIFGSAHKEERFLFPYPIAFFAALALDATNR---LCLKK 397  
 Query: 401 YRLEHYTVTSNWLALGTVFLFGLLSFSRSVALFRGYHGPLDLYPEFYRIATDPTIHTVPE 460  
 ++ N L++ + F +LS SR+ ++ Y +++Y T+ T +  
 Sbjct: 398 LGMD-----NLSILFILCFAILSASRTYSIHNNYGSVHEIYRSLNAELTNRT-NFKNF 450  
 Query: 461 GRPVNVCVGKEWYRFPSSFLPDNW-----QLQFIPSEFRGQLPKPFAEGPL---ATRI 511  
 P+ VCVGKEW+RFPSSF +P +++FI SEFRG LPKPF + TR  
 Sbjct: 451 HDPIRVCVGKEWHRFPSSFFIPQTVSDGKKVEMRFIQSEFRGLLPKPFLLKSDKLVEVTRH 510  
 Query: 512 VPTDMNDQNLLEEPSRYIDISKCHYLVDLDTMRETREPKEYSSNKEEW 558  
 +PT+MN+ N EE SRY+D+ C Y+VD+D M ++ REP + ++ +  
 Sbjct: 511 IPTEMNNLNQEEISRYVDLSDCDYVVDVD-MPQSDREPDRKMRQNY 556

Pedant information for DKFZphutel\_20m24, frame 2

#### Report for DKFZphutel\_20m24.2

[LENGTH] 611  
 [MW] 69863.78  
 [pI] 8.91  
 [HOMOL] SWISSPROT:YTH3\_CAEEL HYPOTHETICAL 75.5 KD PROTEIN C14A4.3 IN CHROMOSOME II. 2e-93  
 [FUNCAT] 09.01 biogenesis of cell wall [S. cerevisiae, YNL219c] 4e-69  
 [FUNCAT] 01.06.01 lipid, fatty-acid and sterol biosynthesis [S. cerevisiae, YNL219c] 4e-69  
 [FUNCAT] 01.05.01 carbohydrate utilization [S. cerevisiae, YNL219c] 4e-69  
 [PIRKW] glycosyltransferase 9e-68  
 [PIRKW] transmembrane protein 9e-68  
 [PIRKW] hexosyltransferase 9e-68  
 [PROSITE] MYRISTYL 9  
 [PROSITE] CAMP\_PHOSPHO\_SITE 1  
 [PROSITE] CK2\_PHOSPHO\_SITE 7  
 [PROSITE] PKC\_PHOSPHO\_SITE 6  
 [PROSITE] ASN\_GLYCOSYLATION 2  
 [KW] TRANSMEMBRANE 7  
 [KW] LOW\_COMPLEXITY 6.71 %

SEQ MASRGARQRLKSGASSGDTAPAADKLELLGSREAGGAEHRTLSGNKAGQVWAPEGST  
 SEG .....  
 PRD cchhhhhhhccccccccccccchhhhhhhccccccccccccccccccccccccccccch  
 MEM .....MMMMMM

SEQ AFKCLLSARLCAALLSNISDCDEFNYWEPHYLYIYGEFQTWEYSPAYAIRSYAYLLLH  
 SEG .....  
 PRD hhhhhhhhhhhhhhhhhhhhhccch  
 MEM MM

SEQ AWPAAFHARILQTNKILVIFYFLRCLLAFVSCICELYFYKAVCKKFLHVSRLMLAFLVLS  
 SEG .....  
 PRD cchhhcc  
 MEM MM

SEQ TGMFCSSAFLPSSFCMYTTLIAMTGWYMDKTSIAVLGVAAGAILGWPFSAALGLPIAFD  
 SEG .....  
 PRD cccccccccccccchhh  
 MEM .....MMMMMMMMMMMMMMMM

SEQ LLVMKRWKSFHWSLMALILFLVPVVVIDSYYYGKLVIAPLNIVLYNVFTPHGPDLYGT  
 SEG .....  
 PRD hhhcc  
 MEM MMMMMMM.MMM

SEQ EPWYFYLINGFLNFNVAFA LALLVPLTSLMEYLLQRFHVQNLGHPYWLTLAPMYIWFII  
 SEG .....  
 PRD cccccccccccccchhh  
 MEM .....MM

Prosites for DKFZphutel1\_20m24.2

(No Pfam data available for DKFZphut1\_20m24.2)

DKFZphutel\_21d15

group: uterus derived

DKFZphutel\_21d15 encodes a novel 191 amino acid protein without similarity to known proteins.

No informative BLAST results; No predictive prosite, pfam or SCOP motive.

The new protein can find application in studying the expression profile of testis-specific genes.

unknown

Sequenced by MediGenomix

Locus: /chromosome="3"

Insert length: 5292 bp

Poly A stretch at pos. 5273, polyadenylation signal at pos. 5252

```
1 CTCCCACTAG TGTATGCCTT AATGGTGCCG CTCTTGTCGG CGTCTACGCT
51 TGGGACCTTG GCTTCTGACT TGGAGAGTGT ACAGCTCTGC CCGACGGCAA
101 CCCAGCTTGG GAAGAGAAGC CCCAGCGTGG GCTGGGGCTC AAGGCGCAGG
151 AAGGCCGAGC CCGGCGCGGA CGCAGGCGGC TCCGGGCGGG CTCAGCACCC
201 CCAGGCACCG TCTCCTAGTG ACCGCGGCGC TCGCGGGCCT GCGGGCCGTT
251 GTCCGGGCGA CTGCGCAGCG CGGGCACCCC CGCGGCCCTT CCCCTGGGCG
301 CGCGCGCGAC CTGGGTGCCA TGGCGGCAGC GCGCGTGACA GGCCAGCGGC
351 CTGAGACCGC GCGGCGCGAG GAGGCCTCGA GGCCGCAGTG GCGCGCGCCA
401 GACCACTGCC AGGCTCAGGC GCGGCGCGGG CTGGGCGACG GCGAGGACGC
451 ACCGGTGCGT CCGCTGTGCA AGCCCCGCGG CATCTGCTCG CGCGCCTACT
501 TCCTGGTGCT GATGGTGTTT GTGCACCTGT ACCTGGGTAA CGTGCTGGCG
551 CTGCTGCTCT TCGTGCACTA CAGCAACGGC GACGAAAGCA GCGATCCCGG
601 GCCCCAACAC CGTGCCCAAG GCCCCGGGCC CGAGCCCACC TTAGGTCCCC
651 TCACCCGGCT GGAGGGCATC AAGGTGAGGA CCTCCCTGCC CCGCCGCGCT
701 CAGGCCCTTG CACGGCTGAG CCCGAGAGGA CCGCGCTCA GCGCGGGTCC
751 CCACGCTGCC CCGGCGCTG CTCTGCGTCG GTCCCGCGCG CTCCCACTCA
801 CTCGCTGCTG GTCGCTCTCC GGGCGGGGCG GACTTGGCCC TTTTGGGCA
851 GCGCGGTCTG GCGCCCCAGC TGCCCGCTGT GCGCCTTTTC CTTAGGTGGG
901 GCACGAGCGT AAGGTCCAGC TGGTCACCGA CAGGGATCAC TTCATCCGAA
951 CCCTCAGCCT CAAGCCGCTG CTCTTCGAAA TCCCCGGCTT CCTGACTGAT
1001 GAAGAGTGTG GGCTCATCAT CCATCTGGCG CAGATGAAGG GGTACAGCG
1051 CAGCCAGATC CTGCTACTG AAGAGTATGA AGAGGCAATG AGCACTATGC
1101 AGGTGAGCCA GCTGGACCTC TTCCGGCTGC TGGACCAGAA CCGTGATGGG
1151 CACCTTCAGC TCCGTGAGGT TCTGGCCAG ACTCGCCTGG GAAATGGATG
1201 GTGGATGACT CCAGAGAGCA TTCAGGAGAT GTACGCCGCG ATCAAGGCTG
1251 ACCCTGATGG TGACGGTGAG CTCACACCTC TGCACAGTCC TATCCCGTG
1301 AGCCTCTCTG CCACTCCAG GTGCACAATT TTGAAAACCT GGGCCCTTCC
1351 CCCACAGCCA GGCAGCCTCT CTGCACCCCT TTATAGTGGC CAGAGATGGG
1401 GAGGTGAAGA TCCAGCCTTG CTTTTTACCC CTGGGAAGTA GGCAGGCAGC
1451 CAGGCCCCCC GTTCCCTTG GTGATGGTCT CGAGGGCAGT TCTTGGAGAC
1501 CTTTTTGATA ACATCAGGCA GAGTTGAGAG CCTGGGGACA GGAAGTAGGG
1551 CTGCTAGTTG GCAGAGAACA GAGTGGGTGG AGCAGGAGCA AGGCGACAGT
1601 GAGGCCAGCT AGAGCTTGGC TGTTTACCTT GCTCCATCCA TCTCTCCAGC
1651 CAGACACGAG GTCCACCCCA GCAGACAGCT TCCCTGGTCT AAGTGAGGTC
1701 TCCCTTGCTT TCCTCTTGTC CACCTGGAGT CATGCCGAAG CGCCTAAAAT
1751 GGTAGTGCTG CTACCTGTGC TAACTGCTGG GGAGGGGTGG GCAGGGAAGC
1801 TGTATGCAA GTGGTGCCCC CTCTGGTAAT AACTCTCAGG AGGTTTCTGA
1851 GGTGTGGTCA TCACCTCAT GCCCAAATTC TGGACCAAGA GAGGAAGATA
1901 CAGCAGTTAG AAAGGACTTG GAACAGTGGC TTTGCGGCTG GTGAACCAGA
1951 GTGAAGAATC TGGCCGTGAC CTGGCTGCCA CACTGCTATA GGCCCCAGAA
2001 CAGAGGTGGT GACAGTCTCA CAGCCCTTGA ATGTCCCCCA CCCTCAGAGG
2051 AATCTGGGCC AAAGAGTGGA AGGTGATGTC CTTGGGTGAG CCAGAATAAC
2101 ATGGAGCAAA GATACCAACT ACTCTCCAG AACCCCAAGA GGGTAGAACC
2151 CCTGCTTAAT GGTGTGAGCA GGGACAGTGG AGAATGTTCT CATGAGAGGG
2201 GGTGGCCTGA CTTTCGTTGC TAAATGGGCT GGTAAACGAG TAGGCAGGGC
2251 TGGCGAAGTA GGTTCACCCC AGGATGAAAC CTGGGGTCAT GAGGAACTCC
2301 CCGGGGGCTG GCCCTGCTTG CACCTGGCG TATGTATGTA AGGCCCTGGA
2351 TGAGGCCAG CACTGCCTGC TCTCTCTCA CCCTCCACAG GCCGGAGAGT
2401 GGCCACCACT CTATATAGCC AGGCTGGAAG GCCAGGGTCC TGGCCATATG
2451 GCTCAAGCTT CTTTGGAGA ACCTTCTCTG GCCACTCTAA TAGGGGGTGG
2501 GCCTCTTTCT TCTTAGGGCC AAATTAGGGC TTAACCTGAG AAAAGGAACT
2551 GCTCTGGGTC TTCCTGTAAG GCCTGATGTG ACAGAAACCA GGTTTCATCTG
2601 ACCCAAAAGT CCAGGTGGGG GACAAGTGTG CAAGGCCCTT CAGTGCCTGA
2651 GCTCAGGGGC TGCTGCTGCC TTTGGGTTAG GTAGGGAAGT GCAGCCTGCC
2701 ACTGTTGCCT CCCAATATGG GCTTGGTGGG CATTGATGGT GGGTGCCCTG
2751 TGCAGGAGTG CTGAGTCTGC AGGAGTTCTC CAACATGGAC CTTCCGGGAGT
2801 TCCACAAGTA CATGAGGAGC CACAAGGCAG AGTCCAGTGA GCTGGTGCGG
```

```

2851 AACAGCCACC ATACCTGGCT CTACCAGGGT GAGGGTGGCC ACCACATCAT
2901 GCGTGCCATC CGCCAGAGGT GAGCACCTGA AGCTGTTCTC ACTGGAGCAG
2951 GGGGAGAAGA CTGGGCAGGG CCTCCACAGA AGTCCTTGTC TGGGGCCAAG
3001 AGGACAGAAT GGATTAACCC ATTTGGGATT AAGTTCCATT TGTAGACCA
3051 GGATTGGGAC CCACTGAAAG ACAGGCAATT AACAAAGGCA AATTAGCCCT
3101 CCTTGCAGGC ACACAATGGG CAACTGGGGT TAGATAGAGA TTGAGCACTT
3151 CTTTCTGATT AGATAAATGA CCTCTTATCT TTGACCCCTT ATCTGACCCC
3201 GTCACAGCAG GAAAAGGGTT TTTAAATAAA CAACTTTCTT CCAGGGAGGA
3251 GGACCTCAGG ACTCCCCGCC CCCTTTATTT AGTGGAAATG TCAACATTTC
3301 CACATAGCAG GTGTCTCTGT CTTTGGCATC TGAGGGAGAA GGATCATCAT
3351 GAGTAACCCC CTCCTGCTCT TACAGGGCCA GTCTGAGATG GCTTAAGGGA
3401 CTTCAGGGG AGGTGGGTAG GGGCAAAGCT TGTGGCAGGC CTAGGGTCCA
3451 CCTTGGCCAG CTCTTCAGA TCACCACCTT GCCTGGGGCT GCCCAGCCAA
3501 ATGCCTCCTG CCCACCAGGG TGCTGCGCCT CACTCGCCTG TCGCCTGAGA
3551 TCGTGGAGCT CAGCGAGCCC CTGCAGGTG TTCGATATGG TGAGGGGGGC
3601 CACTACCATG CCCACGTGGA CAGTGGGGCT GTGTACCCAG AGACCATCTG
3651 CTCCCATACC AAGCTGGTAG CCAACGAGTC TGTACCCTTC GAGACCTCCT
3701 GCCGGCAAGT ATCTCCCAAC TGGGGGCTGC CTTCAATCCT CAGACCAGGA
3751 ACACCCATGA CACAGGCACA GCCCTGCACT GTGGGCGTGC CCCTTGGCAT
3801 GGGGCCAGGA GATCACTGGG TTATCCCGGT TAGTGATGCC CTCACCTCTC
3851 CCCACAAGTT GTTTACCCAA TGGCTGGAAA GGGGTGGGTA CTGGTCATCG
3901 TGACCACTGG AGTCAACACA GACTGATGTA CCCACAGACA CCAAACTTG
3951 CCCCTGAGT TCTGAAGCAA GGGGCAAGGC TGGGCCCTTA GCTTGTCTG
4001 CCCATTCTCT CAGGTGTTGA TCTTGATTCC ACTTAGAGAA GCTGAAGCTG
4051 TGCTCCCTCT CCCTGTCAAG CCAGTCTCTT CCTCTTCAGG TGGCTGTTCT
4101 GGGCCAGCCC CTTCCCATCC CCAAGGAGCC CTTCAGCGCG CCCTGTTGCT
4151 TCTGCTAGCC TACCTTTCCC TGCCAGGCC TTGCTCAGGG CCATGGCATT
4201 TAACCTAAGT CACCTGTGAT CTTGGCCAAA AAACCATTGC AACTCACAGT
4251 AAGAGACTGG GTTTCGGGGA AGGAGGGGCT AGGGACATTT TGGCACTGGC
4301 CTGCCCTATT GTCTCCATC CTAGTCTGTC CTGGTCCCTG GCAACAGGAA
4351 CTTGGGCAGC TTATCTGCCC CACAGGTAAG CCCCTGGGAG CATCCACAAC
4401 TGGGGACCTG CTCAGTGCCC CCCCTGCCTT ACAGCTACAT GACAGTGCTG
4451 TTTTATTGTA ACAACGTCAC TGGTGGGGGC GAGACTGTTT TCCCTGTAGC
4501 AGATAACAGA ACCTACGATG AAATGGTAAG GGTCAACTGG GCTATTACTC
4551 TTTTGGGCTG GCAGGGGCTT AGACAAGTGA AGTACACACC TCTCCAGGTC
4601 TAAGGATGTG GGCCCAAATT ATTCTTGGG CATATCTGGT TGGTTTCCCT
4651 TTGGTCAACC TTGGCTGGCC TGGCCATAGA GTGGGGACAG GTTGAACACC
4701 CCACCAACCT GCTGCCACA GAGTCTGATT CAGGATGACG TGGACCTCCG
4751 TGACACACGG AGGCACTGTG ACAAGGAAA CCTGCGTGTC AAGCCCCAAC
4801 AGGGCACAGC AGTCTTCTGG TACAACCTACC TGCCTGATGG GCAAGGTTGG
4851 GTGGGTGACG TAGACGACTA CTCGCTGCAC GGGGGCTGCC TGGTCACGCG
4901 CGGCACCAAG TGATTGCCA ACAACTGCAT TAATGTGGAC CCCAGCCGAG
4951 CGCGGCAAGC GCTGTTCCAA CAGGAGATGG CCCGCTTGC CCGAGAAGGG
5001 GGCACCGACT CACAGCCCGA GTGGGCTCTG GACCGGGCCT ACCGCGATGC
5051 GCGCGTGGAA CTCTGAGGGA AGAGTTAGCC CCGGTTCCCA GCCGCGGGTC
5101 GCCAGTTGCC CAAGATCAGG GGTCCGGCTG TCCTTCTGTC CTGCTGCAGA
5151 CTAAGGTCTT GGCCAATGTC TTGCCCCACC CCGCCAGCCG CGATACGGCG
5201 CAGTTCTTAT ATTCAATGTA TTTATTGTGT ACTGACTCCA TCTGCCCCGT
5251 CAAATAAAAA ACCACAAGGT TCGAAAAAAA AAAAAAAGG GG

```

## BLAST Results

Entry HSU64252 from database EMBL:  
 Human STS sequence NOTI-225.  
 Score = 959, P = 1.2e-36, identities = 195/199

## Medline entries

No Medline entry

## Peptide information for frame 1

ORF from the beginning to 351 bp; peptide length: 118  
 Category: questionable ORF  
 Classification: no clue

```

1 LPLVYALMVP LLSASTLGLT ASDLESVQLC PTATQLGKRS PSVWGSRRR
51 KAEPGADAGG SGRAQHPQAP SPSDRGARGP GGRCPGDCAA RAPPRPLPWA
101 RARPGCHGGS GGDRPAA

```

## BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphut1\_21d15, frame 1

No Alert BLASTP hits found

#### Peptide information for frame 2

ORF from 320 bp to 892 bp; peptide length: 191

Category: putative protein

Classification: no clue

```

1 MAAAVTQGR PETAAEEAS RPQWAPPDHC QAQAAAGLGD GEDAPVRPLC
51 KPRGICSRAY FLVLMVFVHL YLGNVLALLL FVHYSNGDES SDPGPQHRAQ
101 GPGPEPTLGP LTRLEGIKVR TSLPRRAPGP ARLSPRGPAL SPGPHAAPGA
151 ALRRSRALPL TRLLSLSGPG RLGPFWAARS GAPAAARCAFF P

```

#### BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphut1\_21d15, frame 2

PIR:EDBE75 immediate-early protein IE175 - human herpesvirus 1, N = 2,  
Score = 106, P = 0.0067

>PIR:EDBE75 immediate-early protein IE175 - human herpesvirus 1  
Length = 1,298

#### HSPs:

Score = 106 (15.9 bits), Expect = 6.7e-03, Sum P(2) = 6.7e-03  
Identities = 36/103 (34%), Positives = 44/103 (42%)

```

Query: 87 GDESSDPGPQHRAQGGPGPEPTLGLPLTRLEGIKVRTSLPRRA-PGPARLS-PRGPALSPGP 144
      G + PGP G GP P P T+ G S R P PA S P GP +P
Sbjct: 726 GRKRKSPGPARPPGGGGPRP---PKTKKSGADAPGSDARAPLPAPAPPSTPPGPEPAPAQ 782

```

```

Query: 145 HAAPGAALRRSRALPLT-RLLSLSGPGRLGPFWAARS GAPAAARCAP 189
      AAP AA ++R P+ GP LG W + P+ AP
Sbjct: 783 PAAPRAAAQARPRPVAVSRRPAEGPDPLGG-WRRQPPGPSHTAAP 827

```

Score = 40 (6.0 bits), Expect = 6.7e-03, Sum P(2) = 6.7e-03  
Identities = 8/21 (38%), Positives = 9/21 (42%)

```

Query: 28 DHCQAQAAAGLGDGEDAPVRP 48
      DH + A G G AP P
Sbjct: 212 DHAREARAVGRGPSSAAPAP 232

```

#### Pedant information for DKFZphut1\_21d15, frame 1

#### Report for DKFZphut1\_21d15.1

```

[LENGTH] 117
[MW] 11797.32
[pI] 10.68
[KW] Irregular
[KW] SIGNAL_PEPTIDE 22
[KW] LOW_COMPLEXITY 38.46 %

```

```

SEQ LPLVYALMVPLLSASTLGLTASDLESVQLCPTATQLGKRSPSVGWSRRRKAEPGADAGG
SEG .....xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx
PRD cccccccccccccccccccccchhhhhhhhhcccccccccccccccccccccccccccccccc

```

```

SEQ SGRAQHPQAPSPSDRGARGPGGRC PGDCAARAPPRLPWARARPGCHGGSGGDRPAA
SEG .....xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx
PRD ccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc

```

(No Prosite data available for DKFZphut1\_21d15.1)

(No Pfam data available for DKFZphute1\_21d15.1)

Pedant information for DKFZphute1\_21d15, frame 2

Report for DKFZphute1\_21d15.2

```
{LENGTH}      191
{MW}           19916.88
{pI}           10.43
{KW}           TRANSMEMBRANE 1
{KW}           LOW_COMPLEXITY 29.84 %

SEQ  MAAAAVTGQRPETAAEEASRPQWAPPDHCQAQAAAGLGDGEDAPVRPLCKPRGICSRAY
SEG  .....
PRD  cccceeeccccchhhhhhhhhccccccchhhhhhhccccccccccccccccccccchhhh
MEM  .....

SEQ  FLVLMVFVHLYLGNVLALLLVHYSNGDESSDPGPQHRAQGPPEPTLGPLTRLEGIKVR
SEG  .....
PRD  hhhhhhhhhhhhhhhhhhhhhccccccccccccccccccccccccccccccccceeeee
MEM  ....MMMMMMMMMMMMMMMM.....

SEQ  TSLPRRAPGPARLSPRGPALSPGPHAAPGAALRRSRALPLTRLLSLSGPGRLGPFWAARS
SEG  .....
PRD  eccccccccccccccccccccccccccccchhhhhhhccccceccccccccchhhhhc
MEM  .....

SEQ  GAPAAARCAPPF
SEG  xxxxxxxxxx..
PRD  cccccccccc
MEM  .....
```

(No Prosite data available for DKFZphute1\_21d15.2)

(No Pfam data available for DKFZphute1\_21d15.2)

DKFZphutel\_22d2

group: signal transduction

DKFZphutel 22d2 encodes a novel 580 amino acid putative GTP-binding protein related to the ras protein. Additionally, the putative protein contains an EF-hand for calcium-binding.

G-proteins are involved in various signal transduction pathways, transferring the signal of a cellular receptor to an intracellular signal cascade.

The new protein can find clinical application in modulating/blocking the response to a cellular receptor.

similarity to GTP-binding proteins

complete cDNA, complete cds, potential start at Bp 64, EST hits  
complete cds according to K08F11.5 and YAL048c

Sequenced by BMFZ

Locus: /map="17"

Insert length: 3247 bp  
Poly A stretch at pos. 3230, no polyadenylation signal found

```

1 CTCCTGGTGA GAGGAGTCCA CTCCGTGCGT GCGGGCGGAG GCCGGCCCCC
51 GAGAGCCGCC GACATGAAGA AAGACGTGCG GATCCTGCTG GTGGGAGAAC
101 CTAGAGTTGG GAAGACATCA CTGATTATGT CTCTGGTCAG TGAAGAATTT
151 CCAGAAAGAGG TTCCTCCCCG GGCAGAAAGAA ATCACCATTTC CAGCTGATGT
201 CACCCCAGAG AGAGTTCCAA CACACATTGT AGATTACTCA GAAGCAGAAC
251 AGAGTGATGA ACAACTTCAT CAAGAAATAT CTCAGGCTAA TGTCTCTGT
301 ATAGTGATAT CCGTTAAACA CAAGCATTCT ATTGATAAGG TAACAAGTCG
351 ATGGATTCCCT CTCATAAATG AAAGAACAGA CAAAGACAGC AGGCTGCCTT
401 TAATATTGGT TGGGAACAAA TCTGATCTGG TGGAATATAG TAGTATGGAG
451 ACCATCCTTC CTATTATGAA CCAGTATACA GAAATAGAAA CCTGTGTGGA
501 GTGTTACGCG AAAAACCTGA AGAACATATC AGAGCTCTTT TATTACGCAC
551 AGAAAGCTGT TCTTCATCCT ACAGGGCCCC TGACTGCCC AGAGGAGAAG
601 GAGATGAAAC CAGCTTGATAT AAAAGCCCTT ACTCGTATAT TTAATAATATC
651 TGATCAAGAT AATGATGGTA CTCTCAATGA TGCTGAACTC AACTTCTTTC
701 AGAGGATTTG TTTCAACACT CCATTAGCTC CTCAGCTCTT GGAGGATGTC
751 AAGAATGTAG TCAGAAAACA TATAAGTGAT GGTGTGGCTG ACAGTGGGTT
801 GACCCCTGAA GGTTTTCTCT TTTTACACAC ACTTTTATC CAGAGAGGGA
851 GACACGAAAC TACTTGGACT GTGCTTCGAC GATTGGTTA TGATGATGAC
901 CTGGATTGTA CACCTGAATA TTTGTTCCCT CTGCTGAAA TACCTCTCTGA
951 TTGCACTACT GAATTAATC ATCATGCATA TTTATTTCTC CAAAGCACCT
1001 TTGACAAGCA TGATTGGGAT AGAGACTGTG CTTTGTCAAC TGATGAGCTT
1051 AAAGATTAT TTAAGTTTT CCCTTACATA CCTTGGGGGC CAGATGTGAA
1101 TAACACAGTT TGTACCAATG AAAGAGGCTG GATAACCTAC CAGGGATTCC
1151 TTTCCCAGTG GACGCTCAGC ACTTATTTAG ATGTACAGCG GTGCTTGGA
1201 TATTTGGGCT ATCTAGGCTA TTCAATATTG ACTGAGCAAG AGTCTCAAGC
1251 TTCAGCTGTT ACAGTGACAA GAGATAAAAA GATAGACCTG CAGAAAAAAC
1301 AAATCTCAAG AAATGTGTTT AGATGTAATG TAATTGGAGT GAAAAACTGT
1351 GGGAAAAGTG GAGTTCTTCA GGCTCTTCTT GGAAGAACT TAATGAGGCA
1401 GAAGAAAATT CGTGAAGATC ATAAATCCTA CTATGCGATT AACACTGTTT
1451 ATGTATATGG ACAAGAGAAA TACTTGTTGT TGCATGATAT CTCAGAATCG
1501 GAATTTCTAA CTGAAGCTGA AATCATTTGT GATGTTGTAT GCCTGGTATA
1551 TGATGTCAGC AATCCCAAAT CCTTTGAATA CTGTGCCAGG ATTTTAAAGC
1601 AACACTTTAT GGACAGCAGA ATACCTTGCT TAATCGTAGC TGCAAAGTCA
1651 GACCTGCATG AAGTTAAACA AGAATACAGT ATTTACCTTA CTGATTCTTG
1701 CAGGAAACAC AAAATGCCTC CACCACAAGC CTTCACTTGC AATACTGCTG
1751 ATGCCCCCAG TAAGGATATC TTTGTTAAAT TGACAACAAT GGCCATGTAT
1801 CCGTAAGTAC TTGCTGTCTT CATTTTCATG TTGATGGTGT CATAACATTG
1851 CATGCCATTA TTAGCCATGA AGGGAATATC TTTGTCACAT AGGAATTGTT
1901 CAGCAACAGA AAGATACTTT GTAATGAGAA GGTACAAATT TGAGTAAATG
1951 CAAGTTTGGT TTGAATGCCA TAATAAAATG ATATAAACAG TGCTTCTGAC
2001 AATATCTGTA TATTTTGTAG CAGGCTGTAA CTATCTTAAT AGAATAGTAC
2051 AATAAAACAC AACCCCCCAC CCAGCATTA AATAAGTTT TACTGGAATA
2101 AAATGGGTTT GGCATCATGT TGTTTTATGC TTATAAGCA TTTTCATATG
2151 AACAGAAAGT TTATATTTTT CTGTTTTTGA CCTTAGGTAT ATGAAGTTTT
2201 CTAATAATTT TTATTAATTT ATGTTGAAAT TGTGGGTATG CTTCAGTTAG
2251 GATATGTCTT TTTTAAGTGC TGTAAGAGT AGTTGTAATT GGAATTTCTA
2301 CTGTATAAAT GTTTTACATT AAGTGTTACG AGCCACAAAT TTCATGTACA
2351 TTATTTATAT ATCTATACAT GCATATGCAC AAGCACATAA CTGTGGTCAT
2401 CTCTGTAGTT TACTAACTGC CTTAAATTTG CATGGTTCTT AATGGCATTG
2451 GCCTCAAGTA GTGTGTTTGT ATAAATCTCT TTTTGTAAAC AATAAGTTTT
2501 TCAGGCAAGT CGTTTCTCAG GACTTTATAG CTTATTCTAC TTATTCTTAT
2551 GTTAGTCTCT AAATTATTTT TCTTCTTATG AAAACTACAG TGTAACACAG

```

```

2601 AGTAATAATC AAACATTGCT ATAAACCAAG AATGACATTT TTCAAAAAGG
2651 TGTGTGATTG TACAGATTTT TAAAGTCAGT TAACTTTACT GCTATTTTAT
2701 TACCTAATAC TTTTTTTAGA TGCAACAAAC CCTTGAATTT CTATTTGTAT
2751 TCGAAGACAA GTCATTCTTA TTATTATAGA ATAACCAAAA CCTTATTTAT
2801 GTTTTACCTT TGCTTTAAAA CTCTCATGTA TGTATCTAC AGAGAGGATC
2851 ATTACAGAGA CAGACTCTCC CGAGACATGG GCCACACTGA TAGAATAGAG
2901 AATTTGAGAA AAATCTGGGT CTTTCTAAAA ACTGCTTTGT AAGTTACTTT
2951 TTCTTTATGA CTTCTGTGGG ATTTTGTGA TATTTCTTA GAGAATGACC
3001 AAATCTCCTT TCTTGCCATA ATTAACATTT AGTAATTATG TAGAAACGCA
3051 CTGCTTGGTC AGGCTTCCTG CCTAGCTATA TATTACGTTG TCTTCCTTAC
3101 TACATAAATG TACTTCTTTA ATCTTGTGAT TACAGTAACT GCAAGTGTGT
3151 TTTTACATCT GCATTTTAA AACATTTTAC TGTAATTCTG TTGTGTGTGT
3201 GTGTGTTATA TGATAAATGT ACATACATGG AAAAAAAAAA AAAAAA

```

## BLAST Results

-----

Entry AC004527 from database EMBL:  
 \*\*\* SEQUENCING IN PROGRESS \*\*\*. NF1-related locus, Direct Submission;  
 HTGS phase 1, 10 unordered pieces.  
 Score = 1899, P = 1.1e-78, identities = 387/396

Entry HSI48355 from database EMBL:  
 human STS SHGC-31220.  
 Score = 1826, P = 7.5e-78, identities = 388/406

## Medline entries

-----

No Medline entry

## Peptide information for frame 1

-----

ORF from 64 bp to 1803 bp; peptide length: 580  
 Category: similarity to known protein

```

1 MKKDVRILLV GEPRVGKTSI IMSLVSEEF EEVPPRAEEI TIPADVTPER
51 VPTHIVDYSE AEQSDQLHQ EISQANVICI VYAVNNKHSI DKVTSRWIPL
101 INERTDKDSR LPLILVGNKS DLVEYSSMET ILPIMNQYTE IETCVCESAK
151 NLKNISELFY YAKKAVLHPT GLYCPEEKE MKPACIKALT RIFKISDQDN
201 DGTLNDALN FFQRICFNTP LAPQALDVK NVVRKHISDG VADSGTLKLG
251 FLFLHTLFIQ RGRHETTWTV LRRFGYDDDL DLTPEYLFPL LKIPPDCTTE
301 LNHAYLFLQ STFDKHLDR DCALSPDELK DLFKVFPYIP WGPDVNNTVC
351 TNERGWITYQ GFLSQWLTIT YLDVQRCLEY LGYLGYSILT EQESQASAVT
401 VTRDKKIDIQ KKQTQRNVFR CNVIGVKNCG KSGVLQALLG RNLMRQKKIR
451 EDHKSYYAIN TVYVYGQEKY LLLHDISESE FLTEAEIICD VVCLVYDVSN
501 PKSFYECARI FKQHFMSRI PCLIVAAKSD LHEVKQEYSI SPTDFCRKHK
551 MPPQAFCTN TADAPSKDIF VKLTTMAMY

```

## BLASTP hits

No BLASTP hits available

## Alert BLASTP hits for DKFZphut1\_22d2, frame 1

TREMBL:CEUK08F11\_3 gene: "K08F11.5"; Caenorhabditis elegans cosmid  
 K08F11., N = 1, Score = 1357, P = 1.1e-138

TREMBL:SPCC320\_4 gene: "SPCC320.04c"; product: "hypothetical protein";  
 S.pombe chromosome III cosmid c320., N = 1, Score = 889, P = 4.4e-89

TREMBL:CEUC47C12\_3 gene: "C47C12.4"; Caenorhabditis elegans cosmid  
 C47C12., N = 2, Score = 408, P = 5.6e-74

PIR:S51971 probable membrane protein YAL048c - yeast (Saccharomyces  
 cerevisiae), N = 1, Score = 677, P = 1.3e-66

>TREMBL:CEUK08F11\_3 gene: "K08F11.5"; Caenorhabditis elegans cosmid  
 K08F11.

Length = 625

HSPs:



Score = 1357 (203.6 bits), Expect = 1.1e-138, P = 1.1e-138  
Identities = 263/582 (45%), Positives = 380/582 (65%)

Query: 4 DVRILLVGEPRVGKTSLSIMSLVSEEFPEEVPPRAEITIPADVTPEVPTTHIVDYSEAEQ 63  
DVRIL+L+G+ GKTSLS+MSL+ +E+ + VP R + + IPADVTPE V T IVD S E+  
Sbjct: 9 DVRIVLIGDEGCGKTSLSVMSLLEDEWVDAVPRRLDRVLIPADVTPEVPTTSIVDLSIKEE 68

Query: 64 SDEQLHQEISQANVICIVYAVNNKHSIDKVTSRWIPLINERTDKDSRLPLILVGNKSDLV 123  
+ + EI QANVIC+VY+V ++ ++D + ++W+PLI + + P+ILVGNKSD  
Sbjct: 69 DENWIVSEIRQANVICVVSVDDESTVDGIQTKWLPILIRQSFGEYHETPVILVGNKSDGT 128

Query: 124 EYSSMETILPIMNQYTEIETCVECSAKNLKNISELFYQAQKAVLHPTGPLYCPEEKEMKP 183  
++ + ILPIM TE+ETCVECSA+ +KN+SE+FYQAQKAV++PT PLY + K++  
Sbjct: 129 A-NNTDKILPIMEANTEVETCVECSARTMKNVSEIFYQAQKAVIYPTRPPLYDADTKQLTD 187

Query: 184 ACIKALTRIFKISDQDNDGTLNDAELNFFQRICFNTPLAPQAEDEVKNVVRKHISDGVAD 243  
KAL R+FKI D+DNDG L+D ELN FQ++CF PL ALEDVK V DGVA+  
Sbjct: 188 RARKALIRVFKICDRDNDGYLSDELNDQKLCFGIPLTSTALEDVKRAVSDGCPDGVAN 247

Query: 244 SGLTLKGFLFLHTLFIQRGRHETTWTVLRFGYDDDLDTPEYLFPLLKIPPDCTTELNH 303  
L L GFL+LH LFI+RGRHETT WLR+FGY+ L L+ +YL+P + IP C+TEL+  
Sbjct: 248 DSLMLAGFLYLHLFIERGRHETTAVLRKFGYETSLKLSEDLYPRITIPVGCSTELSP 307

Query: 304 HAYLFLQSTFDKHDLDRCALSPDELKDLFKVFPYIPWGPVNNVTCTNERGWITYQGFL 363  
F+ + F+K+D D+D LSP EL++LF V P D + TN+RGW+TY G++  
Sbjct: 308 EGVQFVSALFEKYDEDKDGLSPSELQNLFSVCPVPVITKDNILAETNQRGWLYTYNGYM 367

Query: 364 SQWTLTTYLDVQRCELYLGYLSILTEQESQAS----AVTVTRDKKIDLQKKQTQRNVF 419  
+ W +TT +++ + E L YLG+ + +A ++ VTR++K DL+ T R VF  
Sbjct: 368 AYNMNTTLINLTQTFEQLAYLGFPVGRSGPGRAGNTLDSIRVTRERKKDLNHGTDKRVF 427

Query: 420 RCNVIGVKNCGSGVLQALLGRNLMRQKKIREDHKSYYAINTVYVYGQEKYLLLDHI--- 476  
+C V+G K+ GK+ +Q+L GR + +I H S + IN V V + KYLL ++  
Sbjct: 428 QCLVVGAADAGKTVMQSLAGRGMAADVAQIGRRH-SPFVINRVRVKEESKYLLREVDVL 486

Query: 477 SESEFLTEAEIICDVCLVYDVSNPKSFEYCARIFKQHFMSRIPCLIVAASDLHEVKQ 536  
S + L E DVV +YD+SNP SF +CA +++++F ++ PC+++A K + EV Q  
Sbjct: 487 SPQDALGSGETSADVVAFLYDISNPDSFAFCATVYQKYFYRTKTPCVMIATKVEREEVDQ 546

Query: 537 EYSISPTDFCRKHKMPPPPQAFCTNTADAPSKDIFVKLTMMAMP 580  
+ + P +FCR+ ++P P F+ S IF +L MA+YP  
Sbjct: 547 RWEVPPPEFCRQFELPKPIKFSTGNIGQSSSPIFEQLAMMAVYP 590

Pedant information for DKFZphutel\_22d2, frame 1

#### Report for DKFZphutel\_22d2.1

[LENGTH]	580
[MW]	66541.61
[pI]	5.56
[HOMOL]	TREMBL:CEUK08F11_3 gene: "K08F11.5"; Caenorhabditis elegans cosmid K08F11. 1e-149
[FUNCAT]	99 unclassified protcins [S. cerevisiae, YAL048c] 5e-81
[FUNCAT]	03.04 budding, cell polarity and filament formation [S. cerevisiae, YKR055w] 3e-11
[FUNCAT]	03.99 other cell growth, cell division and dna synthesis activities [S. cerevisiae, YNL098c] 8e-09
[FUNCAT]	10.04.07 g-proteins [S. cerevisiae, YNL098c] 8e-09
[FUNCAT]	03.10 sporulation and germination [S. cerevisiae, YNL098c] 8e-09
[FUNCAT]	11.01 stress response [S. cerevisiae, YNL098c] 8e-09
[FUNCAT]	03.22 cell cycle control and mitosis [S. cerevisiae, YNL098c] 8e-09
[FUNCAT]	01.03.13 regulation of nucleotide metabolism [S. cerevisiae, YNL098c] 8e-09
[FUNCAT]	01.05.04 regulation of carbohydrate utilization [S. cerevisiae, YNL098c] 8e-09
[FUNCAT]	30.03 organization of cytoplasm [S. cerevisiae, YOR101w] 4e-08
[FUNCAT]	11.10 cell death [S. cerevisiae, YOR101w] 4e-08
[FUNCAT]	10.02.07 g-proteins [S. cerevisiae, YPR165w] 7e-08
[FUNCAT]	30.04 organization of cytoskeleton [S. cerevisiae, YPR165w] 7e-08
[FUNCAT]	30.08 organization of golgi [S. cerevisiae, YPR165w] 7e-08
[FUNCAT]	08.07 vesicular transport (golgi network, etc.) [S. cerevisiae, YFL005w] 9e-08
[FUNCAT]	30.09 organization of intracellular transport vesicles [S. cerevisiae, YFL005w] 9e-08
[FUNCAT]	30.02 organization of plasma membrane [S. cerevisiae, YFL005w] 9e-08
[FUNCAT]	08.13 vacuolar transport [S. cerevisiae, YNL093w] 1e-07

```

[FUNCAT]      06.04 protein targeting, sorting and translocation [S. cerevisiae, YNL093w]
1e-07
[FUNCAT]      08.19 cellular import [S. cerevisiae, YNL093w] 1e-07
[FUNCAT]      10.05.07 g-proteins [S. cerevisiae, YLR229c] 8e-07
[FUNCAT]      03.07 pheromone response, mating-type determination, sex-specific proteins
[S. cerevisiae, YLR229c] 8e-07
[FUNCAT]      10.99 other signal-transduction activities [S. cerevisiae, YCR027c] 3e-06
[FUNCAT]      09.09 biogenesis of intracellular transport vesicles [S. cerevisiae,
YGL210w] 9e-04
[BLOCKS]      BL00410A Dynamin family proteins
[SCOP]         d1plk_ 3.25.1.3.1 cH-p21 Ras protein [human (Homo sapiens) 2e-42
[SCOP]         d1guua_ 3.25.1.3.10 Rap1A [Human (Homo sapiens) 5e-59
[PIRKW]        transmembrane protein 1e-79
[PIRKW]        membrane trafficking 2e-06
[PIRKW]        acetylated amino end 3e-09
[PIRKW]        prenylated cysteine 3e-09
[PIRKW]        signal transduction 1e-07
[PIRKW]        transforming protein 3e-09
[PIRKW]        immediate-early protein 8e-06
[PIRKW]        alternative splicing 4e-08
[PIRKW]        P-loop 1e-10
[PIRKW]        lipoprotein 7e-10
[PIRKW]        proto-oncogene 3e-09
[PIRKW]        methylated carboxyl end 3e-09
[PIRKW]        membrane protein 3e-09
[PIRKW]        GTP binding 1e-10
[PIRKW]        thiolester bond 7e-10
[SUPFAM]       ras transforming protein 1e-10
[PROSITE]      ATP_GTP_A 2
[PROSITE]      MYRISTYL 3
[PROSITE]      EF_HAND 1
[PROSITE]      CAMP_PHOSPHO_SITE 1
[PROSITE]      CK2_PHOSPHO_SITE 14
[PROSITE]      TYR_PHOSPHO_SITE 4
[PROSITE]      PKC_PHOSPHO_SITE 5
[PROSITE]      ASN_GLYCOSYLATION 3
[PFAM]         Ras family (contains ATP/GTP binding P-loop)
[KW]           Irregular
[KW]           3D

```

```

SEQ      MKKDVRILLVGEPRVGKTSLIMSLVSEEFPEEVPPRAEETIPADVTPEVPVTHIVDYSE
1jai-    ...EEEEEEETTTCCHHHHHHHHHHCCCCCCCCCEEEEEETEEEEEEEEEECC
SEQ      AEQSDQLHQEISQANVICIVAVNNKHSIDKVTSRWIPLINERTOKDSRLPLILVGNKS
1jai-    CGGGHHHHHHHHHHHTTEEEETTTTHHHHHHH-HHHHHHHHHHCTTT-TCEEEETTT
SEQ      DLVEYSSMETILPIMNQYTEIETCVECSAKNLKNISELFYYAQAVLHPTGLYPCEEKE
1jai-    TTTTTTTTHHHHHHHHHHHHCCCE-EECTTTTTTHHHHHH.....
SEQ      MKPACIKALTRIFKISDQDNDGTLNDAELNFFQRICFNTPLAPQALEDVKNVVRKHISDG
1jai-    .....
SEQ      VADSGTLTKGFLFLHTLFTIQRGRHETTTVLRFRGYDDDLTPPEYLFPLLKIPDCTTE
1jai-    .....
SEQ      LNHAYLFLQSTFDKHDLDLDRDCALSPDELKDLFKVFPYIPWGPDVNNTVCTNERGWITYQ
1jai-    .....
SEQ      GFLSQWTLTTYLDVQRCLEYLGYLGYSILTEQESQASAVTVTRDKIDLQKKQTQRNVFR
1jai-    .....
SEQ      CNVIGVKNCGKSGVLQALLGRNLMRQKKIREDHKSYYAINTVYVYGQEKYLLHDISESE
1jai-    .....
SEQ      FLTEAEIICDVCLVDVSNPKSFEYCARIFKQHFMDSRIPCLIVAAKSDLHEVKQEYSI
1jai-    .....
SEQ      SPTDFCRKHKMPPOAFTCNTADAPSKDIFVKLTMMAMP
1jai-    .....

```

## Prosites for DKF2phut1\_22d2.1

PS00001	118->122	ASN_GLYCOSYLATION	PDOC00001
PS00001	154->158	ASN_GLYCOSYLATION	PDOC00001
PS00001	346->350	ASN_GLYCOSYLATION	PDOC00001
PS00004	411->415	CAMP_PHOSPHO_SITE	PDOC00004
PS00005	94->97	PKC_PHOSPHO_SITE	PDOC00005
PS00005	105->108	PKC_PHOSPHO_SITE	PDOC00005

PS00005	148->151	PKC_PHOSPHO_SITE	PDOC00005
PS00005	247->250	PKC_PHOSPHO_SITE	PDOC00005
PS00005	414->417	PKC_PHOSPHO_SITE	PDOC00005
PS00006	59->63	CK2_PHOSPHO_SITE	PDOC00006
PS00006	105->109	CK2_PHOSPHO_SITE	PDOC00006
PS00006	126->130	CK2_PHOSPHO_SITE	PDOC00006
PS00006	139->143	CK2_PHOSPHO_SITE	PDOC00006
PS00006	143->147	CK2_PHOSPHO_SITE	PDOC00006
PS00006	196->200	CK2_PHOSPHO_SITE	PDOC00006
PS00006	203->207	CK2_PHOSPHO_SITE	PDOC00006
PS00006	311->315	CK2_PHOSPHO_SITE	PDOC00006
PS00006	325->329	CK2_PHOSPHO_SITE	PDOC00006
PS00006	370->374	CK2_PHOSPHO_SITE	PDOC00006
PS00006	390->394	CK2_PHOSPHO_SITE	PDOC00006
PS00006	477->481	CK2_PHOSPHO_SITE	PDOC00006
PS00006	483->487	CK2_PHOSPHO_SITE	PDOC00006
PS00006	541->545	CK2_PHOSPHO_SITE	PDOC00006
PS00007	153->161	TYR_PHOSPHO_SITE	PDOC00007
PS00007	376->384	TYR_PHOSPHO_SITE	PDOC00007
PS00007	153->162	TYR_PHOSPHO_SITE	PDOC00007
PS00007	448->457	TYR_PHOSPHO_SITE	PDOC00007
PS00008	240->246	MYRISTYL	PDOC00008
PS00008	425->431	MYRISTYL	PDOC00008
PS00008	433->439	MYRISTYL	PDOC00008
PS00017	11->19	ATP_GTP_A	PDOC00017
PS00017	425->433	ATP_GTP_A	PDOC00017
PS00018	197->210	EF_HAND	PDOC00018

## Pfam for DKF2phute1\_22d2.1

HMM_NAME	Ras family (contains ATP/GTP binding P-loop)		
HMM	*KLVLIGDSGVGKSCLLIRFTQNeFnEeYIPTIGvDFYtKTIEIDGKtIK		
	++L+G+ V GK++L ++ EF+EE +P ++ T ++ +++		
Query	6	RILLVGEPRVGKTSLIMSLVSEEFPEE-VPPR-AEEITIPADVTPE RVP	52
HMM	LQIWDTAGQERYRSMRPYYRGAMGFM LVYDITNRqSFENIr.NWweEIr		
	I D E+ + + +A+++ +VY+++N+ S ++++ +W++ I+		
Query	53	THIVDYSEAEQSDEQLHQEISQANVICIVYAVNNKHSIDKVTSRWIPLIN	102
HMM	RHCDrDENVPIMLVGNKCDLEDQRQVStEEGQeFAREWGAIPFMETSAKT		
	+ D+D+ P +LVGNK+DL + ++T + +E+SAK+		
Query	103	ERTDKDSRLPLILVGNKSOLVEYSSMETILPIMNQYTEI-ETCVECSAKN	151
HMM	NiNVEEAFMEIvReIlqrMqeqNqteNinidQpsrnrkrCCCIM*		
	N+ E F+ + +++L + . +++ +++++ + C+		
Query	152	LKNISELFYYAQKAVLHPT-----GLYCPEEKEMK-PACI--	186

DKFZphute1\_22e12

group: signal transduction

DKFZphute1\_22e12 encodes a novel 92 amino acid protein, with similarity to yeast, C.elegans, Drosophila and mammalian proteins.

The Drosophila cni and mammalian cornichon proteins are part of a signal transduction pathway involving the EGF-receptor.

The new protein can find application in modulating the cornichon modulated signal transduction way and also the EGF receptor signaling processes.

strong similarity to S.cerevisiae YGL054c and cornichon

complete cDNA, complete cds, EST hits  
cornichon is required for signal transduction in the EGF-receptor  
signal processing

Sequenced by BMFZ

Locus: unknown

Insert length: 519 bp

Poly A stretch at pos. 499, no polyadenylation signal found

```
1  GTCGGGGCAT CCGAGCGGGT TTGACGGAAG GAGCGGCGGC GACGGAGGAG
51 GAGGATGGAG GCGGTGGTGT TCGTCTTCTC TCTCCTCGAT TGTTGCGCGC
101 TCATCTTCCT CTCGGTCTAC TTCATAATTA CATTGTCTGA TTTAGAATGT
151 GATTACATTA ATGCTAGATC ATGTTGCTCA AAATTAAACA AGTGGGTAAT
201 TCCAGAATTG ATTGGCCATA CCATTGTCAC TGTATTACTG CTCATGTCAT
251 TGCACGTGGT CATCTTCCTT CTCAACTTAC CTGTTGCCAC TTGGAATATA
301 TATCGTATGA TCTTAGCTTT GATAAATGAC TGAAGCTGGA GAAGCCGTGG
351 TTGAAGTCAG CCTACACTAC AGTGCACAGT TGAGGAGCCA GAGACTTCTT
401 AAATCATCCT TAGAACCGTG ACCATAGCAG TATATATTTT CCTCTGGAA
451 CAAAAAATA TTTTGCTGT ATTTTACCA TATAAAGTAT TAAAAAACA
501 TGAACAAAAA AAAAAA
```

## BLAST Results

No BLAST result

## Medline entries

95300228:  
cornichon and the EGF receptor signaling process are necessary for both  
anterior-posterior  
and dorsal-ventral pattern formation in Drosophila.

## Peptide information for frame 1

ORF from 55 bp to 330 bp; peptide length: 92  
Category: strong similarity to known protein

```
1  MEAVVFVFSL LDCCALIFLS VYFIITLSDL ECDYINARSC CSKLNKWWIP
51  ELIGHTIVTV LLLMSLHWFI FLLNLPVATW NIYRMILALI ND
```

## BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphute1\_22e12, frame 1

PIR:S64058 probable membrane protein YGL054c - yeast (Saccharomyces cerevisiae), N = 2, Score = 185, P = 5.7e-17

TREMBL:SPAC2C4\_5 gene: "SPAC2C4.05"; product: "cornichon homolog";

S.pombe chromosome I cosmid c2C4., N = 1, Score = 163, P = 3.7e-12

PIR:S46084 probable membrane protein YBR210w - yeast (Saccharomyces cerevisiae), N = 1, Score = 162, P = 4.8e-12

TREMBL:AF104398.1 product: "cornichon"; Homo sapiens cornichon mRNA, complete cds., N = 1, Score = 141, P = 8e-10

SWISSPROT:CNI\_DROVI CORNICHON PROTEIN., N = 1, Score = 139, P = 1.3e-09

>PIR:S64058 probable membrane protein YGL054c - yeast (Saccharomyces cerevisiae)  
Length = 138

## HSPs:

Score = 185 (27.8 bits), Expect = 5.7e-17, Sum P(2) = 5.7e-17  
Identities = 35/85 (41%), Positives = 56/85 (65%)

Query: 1 MEAVVFVFSLLDCCALIFLSVYFIITLSDLECDYINARSCCSKLNKWWIPELIGHTIVTV 60  
M A +F+ +++ C +F V+F I +DLE DYIN CSK+NK + PE H +++  
Sbjct: 1 MGAWLFILAVVNCINLFGQVHFTILYADLEADYINFIELCSKVNKLITPEAALHGALSL 60

Query: 61 LLLMSLHWFIPLLNPVATWNIYRM 85  
L L++ +WF+FLNLPV +N+ ++  
Sbjct: 61 LFLNGYWFVFLNLPVLAYNLNKI 85

Score = 37 (5.6 bits), Expect = 5.7e-17, Sum P(2) = 5.7e-17  
Identities = 7/9 (77%), Positives = 9/9 (100%)

Query: 82 IYRMILALI 90  
+YRMI+ALI  
Sbjct: 123 LYRMIMALI 131

## Pedant information for DKFZphut1\_22e12, frame 1

## Report for DKFZphut1\_22e12.1

[LENGTH] 92  
[MW] 10614.98  
[pI] 5.04  
[HOMOL] PIR:S64058 probable membrane protein YGL054c - yeast (Saccharomyces cerevisiae)  
5e-14  
[FUNCAT] 03.04 budding, cell polarity and filament formation (S. cerevisiae, YGL054c)  
2e-15  
[PIRKW] transmembrane protein 2e-11  
[PROSITE] CK2\_PHOSPHO\_SITE 3  
[KW] SIGNAL PEPTIDE 33  
[KW] TRANSMEMBRANE 2

SEQ MEAVVFVFSLLDCCALIFLSVYFIITLSDLECDYINARSCCSKLNKWWIPELIGHTIVTV  
PRD ccchhhhhhhhhhhhhhhhhhhheeeccccccccccccccccceehhhhhhhhhhhhh  
MEM .....MMMMMMMMMM

SEQ LLLMSLHWFIPLLNPVATWNIYRMILALIND  
PRD hhhhhhhheeeccccchhhhhhhhhhhccccc  
MEM MMMMMMMMMMMMMMMMMMMMMMMMMMMMM

## Prosite for DKFZphut1\_22e12.1

PS00006	9->13	CK2_PHOSPHO_SITE	PDOC00006
PS00006	26->30	CK2_PHOSPHO_SITE	PDOC00006
PS00006	28->32	CK2_PHOSPHO_SITE	PDOC00006

(No Pfam data available for DKFZphut1\_22e12.1)

DKFZphut1\_22n2

group: uterus derived

DKFZphut1\_22n2 encodes a novel 304 amino acid protein without similarity to known proteins.

No informative BLAST results; No predictive prosite, pfam or SCOP motif.

The new protein can find application in studying the expression profile of uterus-specific genes.

unknown

complete cDNA, complete cds, EST hits

Sequenced by BMFZ

Locus: /map="553.3 cR from top of Chr11 linkage group"

Insert length: 1556 bp

Poly A stretch at pos. 1534, no polyadenylation signal found

```
1 ACAACAGGCT GGTGCTTGG CGTGAATCC TAAAGTGGCC TGGCTTGGAG
51 ACTGGAGTGA GACCCAGGCC CTAGGCTGGG GTTCTTTCCA TTATAGAGGA
101 GACGGATTCA GAAGGGCTAC AGACCAAGGT TGTGAAAAC CAGACATATG
151 ATGAGCGTCT AGAGATTAACT GACTCCGAAG AGTTGCAAG TATTTATCT
201 CCAACCCCAA GACACCAAGG ACTTCCCTCGT TCTGCCCATC TTCCTAACAA
251 GGCTATGGCT GATAACAGCA GTGATGAGTG TGAAGAGGAA AATAACAAGG
301 AGAAGAAGAA GACCTCACAG TTGACACCTC AACGGGGCTT TAGTGAAAAAT
351 GAGGATGACG ATGATGATGA TGATGATTCA TCTGAACTG ATTCTGATTC
401 TGATGATGAT GATGAAGAGC ATGGAGCCCC TCTGGAAGGG GCCTATGACC
451 CTGCAGACTA TGAGCATTTG CCAGTTTCTG CTGAAATTAA GGAACCTCTC
501 CAGTACATCA GTAGGTACAC ACCTCAGTTG ATTGACCTGG ACCACAACT
551 GAAGCCTTTC ATTCTGATT TTATCCCAGC TGTGGGGGAT ATTGATGCAT
601 TCTTAAAGGT CCCACGTCCT GATGAAAAGC CTGACAACCT TGGCCTATTG
651 GTATTGGATG AACCTTCTAC AAAGCAGTCA GACCCACGG TGCTCTCACT
701 CTGGTTAACA GAGAATTCTA AGCAGACAAA CATCACAAA CATATGAAAG
751 TAAAAGCCTC AGAAGATGCA GAAAAGATC CCAAAGCCAT TGACACGTGG
801 ATTGAGAGCA TCTCTGAATT ACACCGTTCT AAGCCCCCTG CGACTGTGCA
851 CTACACCAGG CCCATGCCCG ACATTGACAC GCTGATGCAG GAATGGTCCC
901 CGGAGTTTGA AGAGCTTTTG GGCAAGGTAA GCCTGCCAC GGCAGAGATT
951 GATTGCAGCC TGGCAGAGTA CATTGACATG ATCTGTGCCA TTCTAGACAT
1001 CCCTGCTTAC AAGAGTCGGA TCCAGTCCCT CCATCTGCTC TTTTCCCTCT
1051 ACTCAGAATT CAAGAACTCA CAGCATTTTA AAGCTCTCGC TGAAGGCAAG
1101 AAAGCATTCA CTCCTTCATC CAATTCCACC TCCCAGCTG GAGACATGGA
1151 GACATTAACC TTCAGCTGAG AACTTCCCA AGCTGCTGTT TCAAGGCTGA
1201 GCTGGCCCCCT CTGCCCCAGC TGAGATGGAC AGATCGTTGT CAGCTACTTG
1251 ATGTCCTTGC CCATGCCACA GCTTGGCTCA GGGGCAGTGC ATGTCTCTGT
1301 GCCCTCTCTG CCAGAGGGCA CAGAACATGT TTGTTTAATG AACCTGCCTG
1351 CCTCAGATTG CTGTCCCCGG GGAGTTAATG CATCTACACC ACTGTGGGGA
1401 TTTGAGTTAT AAGAATTGGA ATTTCTGAGA TCCCATGGAG GTTAGATTGG
1451 GAGGAAAGCT TAAAAGATGT CCTTTTGTG AGAGGGATGG AATTGTTTTT
1501 TTTCAATCGT AAAGTTAGTG AGTAAAGATT TTATAAATCA AAAAAAAAAA
1551 AAAAAA
```

## BLAST Results

Entry HS188252 from database EMBL:  
human STS WI-12265.  
Score = 2554, P = 4.1e-109, identities = 556/587

## Medline entries

No Medline entry

## Peptide information for frame 3

ORF from 255 bp to 1166 bp; peptide length: 304  
Category: putative protein

```

1  MADNSSDECE EENNKEKKKT SQLTPQRGFS ENEDDDDDDD DSSETDSDSD
51  DDDEEHGAPL EGAYDPADYE HLPVSAEIKE LFQYISRYTP QLIDLHKLK
101 PFIPDFIPAV GIDDAFLKVP RPDGKPDNLG LLVLDEPSTK QSDPTVLSLW
151 LTENSKQHNI TQHMVKVSL EDAKNPKAID TWIESISELH RSKPPATVHY
201 TRMPDIDTL MQEWSPEFEE LLGKVSLEPTA EIDCSLAEYI DMICAILDIP
251 VYKSRIQSLH LLFSLYSEFK NSQHFALAE GKKAFTPSSN STSQAGDMET
301 LTFS

```

## BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphut1\_22n2, frame 3

PIR:S38149 SIS2 protein - yeast (*Saccharomyces cerevisiae*), N = 1,  
Score = 132, P = 1e-05

>PIR:S38149 SIS2 protein - yeast (*Saccharomyces cerevisiae*)  
Length = 562

## HSPs:

Score = 132 (19.8 bits), Expect = 1.0e-05, P = 1.0e-05  
Identities = 24/63 (38%), Positives = 35/63 (55%)

```

Query:   3  DNSSDECEEEENNKEKKKTSQLTPQRGFSENEEDDDDDDDSDSETDSDSDDEEHGAPLEG 62
          +  DE EEE++ E++ T          +DDDDDDDD + D D DDD++E A  G
Sbjct:  497 EEDDDDEDEEEDDEEDTEDKNENNNDDDDDDDDDDDDDDDDDDEDEDEAETPG 556

```

```

Query:   63 AYD 65
          D
Sbjct:  557 IID 559

```

Score = 122 (18.3 bits), Expect = 1.4e-04, P = 1.4e-04  
Identities = 20/52 (38%), Positives = 33/52 (63%)

```

Query:   4  NSSDECEEEENNKEKKKTSQLTPQRGFSENEEDDDDDDDSDSETDSDSDDEE 55
          N+ +E ++E+ +E + T + + N+DDDDDDDD + D D DDD++
Sbjct:  494 NNEEEDDDDEEEDDEEDTEDKNENNNDDDDDDDDDDDDDDDDDDDDD 545

```

Pedant information for DKFZphut1\_22n2, frame 3

## Report for DKFZphut1\_22n2.3

```

[LENGTH]      304
[MW]           34285.85
[pI]           4.37
[PROSITE]      AMIDATION      1
[PROSITE]      CAMP_PHOSPHO_SITE  2
[PROSITE]      CK2_PHOSPHO_SITE  10
[PROSITE]      PKC_PHOSPHO_SITE   1
[PROSITE]      ASN_GLYCOSYLATION  3
[KW]           All_Alpha
[KW]           LOW_COMPLEXITY    11.84 %

```

```

SEQ  MADNSSDECEEEENNKEKKKTSQLTPQRGFSENEEDDDDDDDSDSETDSDSDDEEHGAPL
SEG  .....XXXXXXXXXX.....XXXXXXXXXXXXXXXXXXXXXXXXXXXXX.....
PRD  cccccchhhhhchhhhhcccccccccccccccccccccccccccccccccccccccc

```

```

SEQ  EGAYDPADYEHLPVSAEIKELFQYISRYTPQLIDLHKLKPFIPDFIPAVGDIDAFKVP
SEG  .....
PRD  cccccccccchhhhhhhhhhhhhcccccccccccccccccccccccccccccccccccc

```

```

SEQ  RPDGKPDNLGLLVLEPSTKQSDPTVLSLWLTENSKQHNIQHMVKVSLDAEKNPKAID
SEG  .....
PRD  cccccccccceccccccccccccchhhhhccccccccccccccccchhhhhhhccccch

```

```

SEQ  TWIESISELHRSKPPATVHYTRMPDIDTLMQEWSPEFEEELGKVSLEPTAEIDCSLAEYI
SEG  .....
PRD  hhhhhhhhhccccccccccccccccchhhhhccccchhhhhccccccccccccchhhhhh

```

```

SEQ  DMICAILDIPVYKSRIQSLHLLFSLYSEFKNSQHFALAEKGKAFTPSSNSTSQAGDMET
SEG  .....

```

```

PRD      hhhhhhccccchhhhhhhhhhhhhhhhhhhhhcchhhhhhhccccccccccccccccccccc
SEQ      LTFS
SEG      ....
PRD      cccc
    
```

Prosite for DKFZphutel\_22n2.3

PS00001	4->8	ASN_GLYCOSYLATION	PDOC00001
PS00001	159->163	ASN_GLYCOSYLATION	PDOC00001
PS00001	290->294	ASN_GLYCOSYLATION	PDOC00001
PS00004	17->21	CAMP_PHOSPHO_SITE	PDOC00004
PS00004	18->22	CAMP_PHOSPHO_SITE	PDOC00004
PS00005	138->141	PKC_PHOSPHO_SITE	PDOC00005
PS00006	5->9	CK2_PHOSPHO_SITE	PDOC00006
PS00006	30->34	CK2_PHOSPHO_SITE	PDOC00006
PS00006	43->47	CK2_PHOSPHO_SITE	PDOC00006
PS00006	45->49	CK2_PHOSPHO_SITE	PDOC00006
PS00006	47->51	CK2_PHOSPHO_SITE	PDOC00006
PS00006	49->53	CK2_PHOSPHO_SITE	PDOC00006
PS00006	168->172	CK2_PHOSPHO_SITE	PDOC00006
PS00006	181->185	CK2_PHOSPHO_SITE	PDOC00006
PS00006	185->189	CK2_PHOSPHO_SITE	PDOC00006
PS00006	235->239	CK2_PHOSPHO_SITE	PDOC00006
PS00009	280->284	AMIDATION	PDOC00009

(No Pfam data available for DKFZphutel\_22n2.3)



DKFZphut1\_22o2  
-----

group: uterus derived

DKFZphut1\_22o2 encodes a novel 537 amino acid protein without similarity to known proteins.

No informative BLAST results; No predictive prosite, pfam or SCOP motife.

The new protein can find application in studying the expression profile of uterus-specific genes.

similarity to S.pombe SPBC3E7.03c

complete cDNA, complete cds, EST hits

Sequenced by BMFZ

Locus: map="llp15.5"

Insert length: 2714 bp

Poly A stretch at pos. 2695, polyadenylation signal at pos. 2677

```
1 GCAGGGCAGC GTGGGGGCTG AGATCGTTTC CTGTTGGAAC TTCTGGCCCA
51 AGAAGCGCGG GTCACAAGGA GAGGGGTCAG TTCGGTTCAG AGCGACTCAG
101 CCCCTCGACT CGGGTCTTAA AACCTCCGAG CCGCCAGTTC TGCTCAGGCC
151 CGCGCCCCCT TAAAGCGCCA CCAGACGCTG CGCCCCGTTA AAGCGCCACC
201 AGACGCCCGG CCCCCTCCCG GCCTCCCCCG CGCGCTGGCG CGGGGCTTTC
251 TGGGCCAGGG CGGGGCCGGC GAACTGCGGC CCGGAACGGC TGAGGAAGGG
301 CCCGTCCCGC CTTCCCGGC GCGCCATGGA GCCCGGGCG GTTGCAAGAG
351 CCGTGGAGAC GGGTGAGGAG GATGTGATTA TGGAAGCTCT GCGGTCATAC
401 AACCAAGGAG ACTCCAGAG CTTACGTTT GATGATGCCC AACAGGAGGA
451 CCGGAAGAGA CTGGCGGAGC TGCTGGTCTC CGTCTGGAA CAGGGCTTGC
501 CACCTCCCA CCGTGTATC TGGGTGAGA GTGTCCGAAT CCTGTCCCGG
551 GACCGCAACT GCCTGGACCC GTTACCAGC CGCCAGAGCC TGCAGGCACT
601 AGCCTGTAT GCTGACATCT CTGTCTCTGA GGGGTCCGTC CCAGAGTCCG
651 CAGACATGGA TGTGTACTG GAGTCCCTCA AGTGCCTGTG CAACCTCGTG
701 CTCAGCAGCC CTGTGGCACA GATGCTGGCA GCAGAGGCC GCCTAGTGGT
751 GAACTCACA GAGCGTGTGG GGCTGTACCG TGAGAGGAGC TTCCCCACG
801 ATGTCCAGTT CTTTGACTTG CGGCTCCTCT TCCTGCTAAC GGCACCTCCG
851 ACCGATGTGC GCCAGCAGCT GTTTCAGGAG CTGAAAGGAG TGCCTGTGCT
901 AACTGACACA CTGGAGCTGA CGCTGGGGT GACTCCTGAA GGAACCCCC
951 CACCCAGCCT CTTCCCTCC CAAGAGACTG AGCGGGCCAT GGAGATCCTC
1001 AAAGTGCTCT TCAACATCAC CCTGGACTCC ATCAAGGGGG AGGTGGACGA
1051 GGAAGAGCCT GCCCTTTACC GACACCTGGG GACCTTCTC CGGCACGTGTG
1101 TGATGATCGC TACTGTGGA GACCGCAGC AGGAGTTCCA CGGCCACGCA
1151 GTGAACCTCC TGGGAACTT GCCCTCAAG TGTCTGGATG TTCTCTCAC
1201 CCTGGAGCCA CATGGAGACT CCACGGAGTT CATGGAGTG AATATGGATG
1251 TGATTCTGTC CCTCTCATC TTCTAGAGA AGCGTTTGCA CAAGACACAC
1301 AGGCTGAAGG AGAGTGTAGC TCCCGTCTG AGCGTGCTGA CTGAATGTGC
1351 CCGGATGCAC CGCCAGCCA GGAAGTTCTT GAAGGCCAG GATGGCCAC
1401 CTCGCCAGGT GCTGCCCTC CTGCGGGATG TGAGGACACG CCCTGAGGTT
1451 GGGGAGATGC TCGGAACAA GCTTGTCCG CTCATGACAC ACCTGGACAC
1501 AGATGTGAAG AGGGTGGCTG CCGAGTCTTT GTTTGTCTG TGCTCTGAGA
1551 GTGTGCCCGG ATTCATCAAG TACACAGGCT ATGGGAATGC TGCTGGCCTT
1601 CTGGCTGCCA GGGGCCCTAT GGCAGGAGGC CGGCCGAGG GCCAGTACTC
1651 AGAGGATGAG GACACAGACA CAGATGAGTA CAAGGAAGCC AAAGCCAGCA
1701 TAAACCCTGT GACCGGGAGG GTGGAGGAGA AGCCGCCTAA CCCTATGGAG
1751 GGCATGACAG AGGAGCAGAA GGAGCAGCAG GCCATGAAGC TGGTGACCAT
1801 GTTTGACAAG CTCTCCAGGA ACAGAGTCAT CCAGCCAATG GGGATGAGTC
1851 CCGGGGGTCA TCTTACGTCC CTGCAAGATG CCATGTGCGA GACTATGGAG
1901 CAGCAGCTCT CCTCGGACCC TGACTCGGAC CCTGACTGAG GATGGCAGCT
1951 CTTCTGTCTC CCCATCAGGA CTGGTGTGTC TTCCAGAGAC TTCTTGGGG
2001 TTGCAACCTG GGAAGCCAC ATCCCACTGG ATCCACACCC GCGCCCACTT
2051 CTCATCTTTA GAAACCCCTT CTCTTGACTC CGTTCTGTT CATGATTTCG
2101 CTCTGGTCCA GTTCTCATC TCTGGACTGC AACGGTCTTC TTGTGCTAGA
2151 ACTCAGGCTC AGCCTCGAAT TCCACAGACG AAGTACTTTC TTTTGTCTGC
2201 GCCAAGAGGA ATGTGTTTCA AAGCTGTGTC CTGAGGGCAG GGCCTACCTG
2251 GGCACACAGA AGAGCATATG GGAGGGCAGG GGTTTGGGTG TGGGTGACAC
2301 CAAAGCAGGC ACCATCTGGG ATTGGCACAC TGGCAGAGCC AGTGTGTTGG
2351 GGTATGTGCT GCACTTCCCA GGGAGAAAAC CTGTGAGAAC TTTCATACG
2401 AGATATATCA AACACACCCT TCCAAGGTAT GTATGCTCTG TTGTTCTCTG
2451 CCTGTCTTCA CTGAGCGCAG GGCTGGAGGC CTCTTAGACA TTCTCTTGG
2501 TCCTCGTTCA GCTGCCCACT GTAGTATCCA CAGTGCCCGA GTTCTCGCTG
2551 GTTTTGGCAA TTAAACCTCC TTCTACTGCG TTTAGACTAC ACTTACAACA
2601 AGGAAATATG CCTCTGTGTG ACCATAGATT GAGATTTATA CCACATACCA
2651 CACATAGCCA CAGAAACATC ATCTTGAAT AAAGAAGAGT TTTGGACAAA
2701 AAAAAAAAAA AAAA
```

## BLAST Results

Entry AF015416 from database EMBL:  
Homo sapiens chromosome 11 from 11p15.5 region, complete sequence.  
Score = 3356, P = 2.0e-144, identities = 672/673

Entry HS263253 from database EMBL:  
human STS SHGC-15914.  
Score = 1143, P = 9.0e-46, identities = 245/255

## Medline entries

No Medline entry

## Peptide information for frame 2

ORF from 326 bp to 1936 bp; peptide length: 537  
Category: similarity to unknown protein

```

1 MEPRVAEAV ETGEEDVIME ALRSYNQEHQ QSFTFDDAQO EDRKRLAELL
51 VSVLEQGLPP SHRVIWLQSV RILSRDRNCL DPFTSRQSLQ ALACYADISV
101 SEGSPVESAD MDVVLESKLC LCNVLSSPV AQMLAAEARL VVKLTERVGL
151 YRERSFPHDV QFFDLRLFL LTALRTDVRQ QLFQELKGVV LTTDTLELTL
201 GVTPEGNPPP TLLPSQETER AMEILKVLFN ITLDSIKGEV DEEDAALYRH
251 LGTLLRHCVI IATAGDRTEE FHGHAVNLLG NLPLKCLDVL LTLEPHGDST
301 EFMGVNMDVI RALLIFLEKR LHKTHRLKES VAPVLSVLTE CARMHRPARK
351 FLKAQGWPPP QVLPPLRDVR TRPEVGEMLR NKLVRMLTHL DTDVSRVAE
401 FLFVLCSESV PRFIKYTYG NAAGLLAARG LMAGGRPEGQ YSEDEDTDT
451 EYKEAKASIN PVTGRVEEKP PNPMEGMTTE QKEHEAMKLV TMFDKLSRN
501 VIQPMGMSPR GHLTSLQDAM CETMEQQLSS DPDSDDP

```

## BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphut1\_22o2, frame 2

TREMBL:SPBC3E7\_3 gene: "SPBC3E7.03c"; product: "hypothetical protein";  
S.pombe chromosome II cosmid c3E7., N = 1, Score = 112, P = 0.0023

>TREMBL:SPBC3E7\_3 gene: "SPBC3E7.03c"; product: "hypothetical protein";  
S.pombe chromosome II cosmid c3E7.  
Length = 362

## HSPs:

Score = 112 (16.8 bits), Expect = 2.3e-03, P = 2.3e-03  
Identities = 71/289 (24%), Positives = 124/289 (42%)

```

Query:  215 SQETERAM-EILKVLFNITLDSIKGEVDEEDAALYRHLGTLRLHRCVMIATAGDRTEEFHG 273
          SQ+ E + EIL++LF I+ S E DE+ L L+ + +
Sbjct:  12 SQQNMVLTIELRLFPISKRSYLKEEDEQKILL-----LVIEIWASSLNNPNPSPLRW 65

Query:  274 HAVN-LLG-NLPLKCLDVLTLLEPHGDSTEFMGVNMDVIRALLIFLEKRLHKTH----RL 327
          HA N LL NL L LD + + T + +I + +LEK L+ +
Sbjct:  66 HATNALLSFNLQLSLDQAIYVSEIACQT----LQSILISREVEYLEKGLNLCFDIAAKY 121

Query:  328 KESVAPVLSVLTECARMHRPARKFLKAQGWPPPQVLPPLRDVTRTP-EVGEMLRNKLVR 386
          + ++ P+L++L + +L P D R + + G+ R L+RL
Sbjct:  122 QNTLPPILAILLSSFFNIQNL-----SMLLFPTNDDRKQSLQKGKSFRCLLRL 173

Query:  387 MT-HLDTDVSRVAEFLFVLCSESVPRFIKYTYGNAAGLLAARGLMAGGRPEGQYS--- 442
          +T + + A L LC + + G G A G+ M P + +
Sbjct:  174 LTIPIVEPIGTYIASLLNELCDGDSQQIARIFGAGYAMGISQHSSETMPFSPLSKAASPV 233

Query:  443 -EDEDTDDEYKEAKASINPVTGRV--EEKPPNPMEGMTTEEQKEHEAMKLVMTFDKLSRN 499
          + + +E +I+P+TG + +E +++E+KE EA +L +F +L +N
Sbjct:  234 FQKNSRGQENTEENLAIDPITGSMCTNRNKSQRLE-LSQEEKEREERLFYLFQRLEKN 292

```

Pedant information for DKFZphutel\_22o2, frame 2

## Report for DKFZphutel 22o2.2

```

[LENGTH] 537
[MW] 60372.53
[pI] 5.20
[BLOCKS] BL004151 Synapsins proteins
[PROSITE] MYRISTYL 4
[PROSITE] CK2_PHOSPHO_SITE 13
[PROSITE] PKC_PHOSPHO_SITE 10
[PROSITE] ASN_GLYCOSYLATION 1
[KW] All_Alpha
[KW] LOW_COMPLEXITY 9.50 %

SEQ MEPRVAEAVETGEEDVIMEALRSYNQEHQSQSFITDDAQQEDRKRLAELLVSVLEQGLPP
SEG .....
PRD ccchhhhhhhhhccchhhhhhhhhccccceecchhhhhhhhhhhhhhhhhhhccccc

SEQ SHRVIWLQSVRIILSRDNRCLDPFTSRQSLQALACYADISVSEGSVPESADMDVVLSESLKC
SEG .....
PRD ceeeeeeccccccccccccccccchhhhhhhhhhhhhcееееccccccccchhhhhhhhhhh

SEQ LCNLVLSSPVAQMLAAEARLVVKLTERVGLYRERSFPHDVQFFDLRLFLLLTALRTDVRQ
SEG .....
PRD hhhhhccccchhhhhhhhhhhhhhhhhhhccccccccccccchhhhhhhhhhhhhhhhhhhhh

SEQ QLFQELKGVRLITDTLELTGLVTPPEGNPPPTLLPSQETERAMEILKVLFNITLDSIKGEV
SEG .....
PRD hhhhhhchhhhhhhhhhhhhccccccccccccccccchhhhhhhhhhhhhhhhhhhhhccccc

SEQ DEEDAALYRHLGTLRLRHCVMIAATAGDRTEEFHGHAVNLLGNLPLKCLDVLTLLEPHGDST
SEG .....
PRD hhhhhhhhhhhhhhhhhhhhhhhccccccccccccceeeccccccccceeeeeeccccccc

SEQ EFMGVNMDVIRALLIFLEKRLHKTHRLKESVAPVLSVLTECARMHRPARKFLKAQGWWPPP
SEG .....
PRD eeehhhhhhhhhhhhhhhhhhhhhhccccceeehhhhhhhhhhchhhhhhhhhhhccccccc

SEQ QVLPPLRDVTRPEVGEMLRNKLVRMLTHLTDVKKRAAEFLVLCSESVPFRIKYTGYG
SEG .....
PRD cccccccccccchhhhhhhhhhhhhhhhhhhccccchhhhhhhhhhhhhhhhhhhccccceeeccc

SEQ NAAGLLAARGLMAGGRPEQGYSEDEDTDTDEYKEAKASINPVTGRVEEKPPNPMEGMTEE
SEG .....
PRD cccccccccccccccccccccccccccccccccchhhhhhhhhhhccccceeeccccccccchhh

SEQ QKEHEAMKLVTFDKLSRNRVIQPMGMSPRGHLSLQDAMCETMEQQLSSDPSDPSDP
SEG .....
PRD hhhhhhhhhhhhhhhhhhhhhhhccccccccccccccccccccchhhhhhhhhhhhhhhhhhhhhccccccc

```

Prosite for DKFZphut1 22o2.2

PS000001	230->234	ASN_GLYCOSYLATION	PDOC000001
PS000005	61->64	PKC_PHOSPHO_SITE	PDOC000005
PS000005	69->72	PKC_PHOSPHO_SITE	PDOC000005
PS000005	84->87	PKC_PHOSPHO_SITE	PDOC000005
PS000005	117->120	PKC_PHOSPHO_SITE	PDOC000005
PS000005	145->148	PKC_PHOSPHO_SITE	PDOC000005
PS000005	218->221	PKC_PHOSPHO_SITE	PDOC000005
PS000005	235->238	PKC_PHOSPHO_SITE	PDOC000005
PS000005	324->327	PKC_PHOSPHO_SITE	PDOC000005
PS000005	463->466	PKC_PHOSPHO_SITE	PDOC000005
PS000005	508->511	PKC_PHOSPHO_SITE	PDOC000005
PS000006	12->16	CK2_PHOSPHO_SITE	PDOC000006
PS000006	34->38	CK2_PHOSPHO_SITE	PDOC000006
PS000006	52->56	CK2_PHOSPHO_SITE	PDOC000006
PS000006	99->103	CK2_PHOSPHO_SITE	PDOC000006
PS000006	104->108	CK2_PHOSPHO_SITE	PDOC000006
PS000006	263->267	CK2_PHOSPHO_SITE	PDOC000006
PS000006	371->375	CK2_PHOSPHO_SITE	PDOC000006

WO 01/12659

PCT/IB00/01496

PS00006	388->392	CK2_PHOSPHO_SITE	PDOC00006
PS00006	442->446	CK2_PHOSPHO_SITE	PDOC00006
PS00006	447->451	CK2_PHOSPHO_SITE	PDOC00006
PS00006	491->495	CK2_PHOSPHO_SITE	PDOC00006
PS00006	515->519	CK2_PHOSPHO_SITE	PDOC00006
PS00006	530->534	CK2_PHOSPHO_SITE	PDOC00006
PS00008	57->63	MYRISTYL	PDOC00008
PS00008	420->426	MYRISTYL	PDOC00008
PS00008	424->430	MYRISTYL	PDOC00008
PS00008	430->436	MYRISTYL	PDOC00008

(No Pfam data available for DKFzphute1\_22o2.2)

DKFZphute1\_23e13

group: metabolism

DKFZphtes3\_15j18 encodes a novel 148 amino acid protein with similarity to 27K heat shock proteins.

The novel protein contains a serine protease of the subtilase family with an aspartic acid-containing active site. Subtilases are an extensive family of serine proteases whose catalytic activity is provided by a charge relay system similar to that of the trypsin family of serine proteases but which evolved by independent convergent evolution. The sequence around the residues involved in the catalytic triad (aspartic acid, serine and histidine) are completely different from that of the analogous residues in the trypsin serine proteases. Thus the novel protein is a new member of this family.

The new protein can find application in modulation of proteinase activity in cells and as a new enzyme for proteomics and biotechnological production processes.

heat shock protein HSP27

strong similarity to heat shock 27K proteins

complete cDNA, complete cds, EST hits

Sequenced by EMBL

Locus: /map="578.9 cR from top of Chr12 linkage group"

Insert length: 1854 bp

Poly A stretch at pos. 1831, polyadenylation signal at pos. 1810

```
1 GGTATTATTA GCTCCTGGCT CCGCTCTAGA CCTCAGCGGT TCTGGCTGCC
51 AGCCTGGGCA GCCTGGGAAG CCTGGGAGGA CGGTGGGCTTG CCGGTCTGTC
101 GTGAGGCACT GCGGACGGGG ACCCTCTGGG ATTCTGCTGG ATCTGCCCGG
151 GGGGTACCTT TTGGGGGCTG GGACCCAGT CGAGGGGACA CAACCGTCCC
201 TGGCAGTGGT TGGTTCTGCT TCTCCCTGCA GAAAAGCAGC ATTTTCGGAA
251 GCTGAAGAAAT AAGCTAGCCC AGCCACACCA CCTTGTGTG TGACCTTGGG
301 CAGGTGGTTC TGTCTCTCTG AGCCTCTGTT TCTCTCTGAG CTGAGCAGCC
351 ACCATGGCTG ACGGTCAGAT GCCCTTCTCC TGCCACTACC CAAGCCGCTT
401 GCGCCGAGAC CCTTCCGGG ACTTCCCTCT CTCTCTCGC CTGCTGGATG
451 ATGGCTTTGG CATGGACCCC TTCCAGACG ACTTGACAGC CTCTTGGCCC
501 GACTGGGCTC TGCTCTGCTT CTCTCTCGCC TGGCCAGGCA CCCTAAGGTC
551 GGGCATGGTG CCCCGGGGCC CCACTGCCAC CGCCAGGTTT GGGGTGCCTG
601 CCGAGGGGAG GACCCCCCA CCTTCCCTG GGGAGCCCTG GAAAGTGTGT
651 GTGAATGTGC ACAGCTTCAA GCCAGAGGAG TTGATGGTGA AGACCAAAAG
701 TGGATACGTG GAGGTGTCTG GCAAACATGA AGAGAAACAG CAAGAAGGTG
751 GCATTGTTTC TAAGAACTTC ACAAAGAAAA TCCAGCTTCC TGCAGAGGTG
801 GATCTGTGTA CAGTATTTGC CTCACCTTCC CCAGAGGGTC TGCTGATCAT
851 CGAAGCTCCC CAGGTCCCTC CTTACTCAAC ATTGGAGAG AGCAGTTTCA
901 ACAACGAGCT TCCCCAGGAC AGCCAGGAAG TCACCTGTAC CTGAGATGCC
951 AGTACTGGCC CATCCTTGTG TTGTCCCCAA CCCTAGGGCT TCTCTGATTC
1001 CAGGATACAT TACTTTAGCT GAACCTCAGT TTAGTGCAAG TAAATGTGTA
1051 GAGGGTGCGG GGGTGAGGAC TGACCACAGA TTCCCTGGAT AGTGTAGTGG
1101 TAGATTTCTC CACAGGATAG CGCAATTGGC AAATCATGCT TGGTTGTGTT
1151 AGGCCAAAAT ACTAGTTTTC CTTTCTTAC CTTTCTATC TTGATGAAAA
1201 TGTGTGCATAT TCTATAGTTG CAAAACACAT AAAAGGGGAC TTAACATTTT
1251 ACGTTGTATC TTACTTGAGC TGAATGCAAG GGTACTTTT CTCTGGGGAC
1301 CTCCCCATC ACCCAGGTTT CTACTCTGGG CTCCCGATTG CCATGGCTCC
1351 CAAACCATGC CGCATGGTTT GGTAAATGAA ACCCAGTAGC TAACCCCACT
1401 GTGCTTCCAC ATGCTTGGCC TAAATGGGT GATATACAGG TCTTATATCC
1451 CCATATGGAA TTTATCCATC AACCACATAA AAACAAACAG TGCTTCTGTC
1501 CCTCTGCCCA GATGTGTCCA GCACGTTCTC AAAGTTTCCA CATTAGCACT
1551 CCCTAAGGAC GCTGGGAGCC TGTCACTTTA TGATCTGACC TAGGTCCCCC
1601 CTTTCTTCTG TCCCCTGTGT TTAAGTCGGG ATTTTACAG AGGGAGCTGT
1651 CTCCAGACAG CTCATCAGG AACCAAGCAA AGGCCAGATA GCCTGACAGA
1701 TAGGCTAGTG GTATTGTGTA TATGGGCGGG ACGTGTGTGT CATTATTATT
1751 TGAGTTATGC TGTGTGTTAG GGGTAAATAA CAGTAAATAA TTAATAATAA
1801 TAATAATAAT AATAAAGGAG CTGACGTTCT TAAAAAAGAA AAAAAAATAA
1851 AAAA
```

## BLAST Results

Entry HS286348 from database EMBL:  
human STS TIGR-A002J47.  
Score = 510, P = 1.2e-16, identities = 102/102

## Medline entries

95394379:  
Cloning and sequencing of a cDNA encoding the canine HSP27 protein.

94110260:  
Physiological and pathological changes in levels of the two  
small stress proteins, HSP27 and alpha B crystallin, in rat  
hindlimb muscles

## Peptide information for frame 3

ORF from 354 bp to 941 bp; peptide length: 196  
Category: strong similarity to known protein  
Prosites motifs: SUBTILASE\_ASP (28-39)

1 MADGQMPFSC HYPSRLRRDP FRDSPLSRL LDDGFGMDPF PDDLTASWPD  
51 WALPRLSSAW PGTLRSGMVP RGPTATARFG VPAEGRTPPP FPGEPWKVCV  
101 NVHSFKPEEL MVKTKDGYVE VSGKHEEKQQ EGGIVSKNFT KKIQLPAEVD  
151 PVTVFASLSP EGLLIIEAPQ VPPYSTFGES SFNNELPQDS QEVTCT

## BLASTP hits

No BLASTP hits available

## Alert BLASTP hits for DKFZphut1\_23e13, frame 3

PIR:JC4244 heat-shock 27K protein - dog, N = 1, Score = 304, P =  
4.3e-27

PIR:JN0924 heat shock 27 protein - rat, N = 1, Score = 301, P = 8.9e-27

TREMBL:MM03561\_1 product: "heat shock protein HSP27"; Mus musculus  
heat shock protein HSP27 internal deletion variant b mRNA, complete  
cds., N = 1, Score = 301, P = 8.9e-27

>PIR:JC4244 heat-shock 27K protein - dog  
Length = 209

## HSPs:

Score = 304 (45.6 bits), Expect = 4.3e-27, P = 4.3e-27  
Identities = 80/182 (43%), Positives = 102/182 (56%)

Query: 1 MADGQMPFSC-HYPSRLRRDPFRD-SPLSSRLDDGFGMDPFDDLTASWPDWALPRLSS 58  
M + ++PFS PS DPFRD P SRL D FG+ P++ W W S  
Sbjct: 1 MTERRVPFSLRSPSW---DPFRDWYPAHSRLFDQAFGLPRLPEE---WAQWFG---HS 50

Query: 59 AWPGLTRSGMVP---RGPTATARFGVPAEGR--TPPPFPG-----EPWKVCVNVHSF 105  
WPG +R +P GP A A PA R + G + W+V ++V+ F  
Sbjct: 51 GWPGYVRP--IPPAVEGPAAAAAAPAYSRLSRQLSSGVSEIRQTADRWVSLDVNHF 108

Query: 106 KPEELMVKTKDGYVEVSGKHEEKQEGGIVSKNFTKKIQLPAEVDVPTVFASLSPEGLLI 165  
PEEL VKTKDG VE++GKHEE+Q E G +S+ T K LP VDP V +SLSPEG L  
Sbjct: 109 APEELTVKTKDGVVEITGKHEERQDEHGYISRRLLTPKYTLPPGVDPTLVSSSLSPEGTLT 168

Query: 166 IEAPQVPPYSTFGE 179  
+EAP P + E  
Sbjct: 169 VEAPMPKPATQSAE 182

## Pedant information for DKFZphut1\_23e13, frame 3

## Report for DKFZphut1\_23e13.3

[LENGTH] 196  
[MW] 21604.37

```

[pI]          5.00
[HOMOL]       PIR:JC4244 heat-shock 27K protein - dog 3e-22
[BLOCKS]      BL01031C
[PIRKW]       blocked amino end 1e-13
[PIRKW]       acetylated amino end 4e-13
[PIRKW]       phosphoprotein 7e-21
[PIRKW]       glycoprotein 2e-11
[PIRKW]       heat shock 7e-21
[PIRKW]       molecular chaperone 4e-13
[PIRKW]       alternative splicing 1e-19
[PIRKW]       eye lens 6e-14
[PIRKW]       stress-induced protein 7e-21
[SUPFAM]      alpha-crystallin 7e-21
[PROSITE]     SUBTILASE_ASP 1
[PROSITE]     MYRISTYL 2
[PROSITE]     CK2_PHOSPHO_SITE 2
[PROSITE]     PKC_PHOSPHO_SITE 6
[PROSITE]     ASN_GLYCOSYLATION 1
[PFAM]        Heat shock hsp20 proteins
[KW]          All_Beta
[KW]          LOW_COMPLEXITY 7.14 %

```

```

SEQ  MADGQMFFSCHYPSRLRRDPFRDSPLSSRLDDGFGMDPFPDDLTA SWPDWALPRLSSAW
SEG  .....xxxxxxxxxxxxxxxx.....
PRD  cccccccccccccccccccccccccchhhhhccccccccccccccccccccccccccccc

SEQ  PGTLRSGMVPFRGPTATARFGVPAEGRTPPPFPGEPWKVCNVVHSFKPEELMVTKDKGYVE
SEG  .....
PRD  cccccccccccccchhhhhhhccccccchhhhhheeeeeccccccccccccccccccc

SEQ  VSGKHEEKQEGGIVSKNFTKKIQLPAEVDPTVTFASLSPEGLLIIEAPQVPPYSTFGES
SEG  .....
PRD  eccchhhhccccccccccccccccccccccccccccccccccccccccccccccccccc

SEQ  SFNNELPQDSQEVCTCT
SEG  .....
PRD  cccccccccccccccc

```

## Prosites for DKFZphut1\_23e13.3

PS00001	138->142	ASN_GLYCOSYLATION	PDOC00001
PS00005	27->30	PKC_PHOSPHO_SITE	PDOC00005
PS00005	63->66	PKC_PHOSPHO_SITE	PDOC00005
PS00005	76->79	PKC_PHOSPHO_SITE	PDOC00005
PS00005	104->107	PKC_PHOSPHO_SITE	PDOC00005
PS00005	122->125	PKC_PHOSPHO_SITE	PDOC00005
PS00005	140->143	PKC_PHOSPHO_SITE	PDOC00005
PS00006	47->51	CK2_PHOSPHO_SITE	PDOC00006
PS00006	176->180	CK2_PHOSPHO_SITE	PDOC00006
PS00008	62->68	MYRISTYL	PDOC00008
PS00008	132->138	MYRISTYL	PDOC00008
PS00136	28->39	SUBTILASE_ASP	PDOC00125

## Pfam for DKFZphut1\_23e13.3

```

HMM_NAME      Heat shock hsp20 proteins
HMM            *AMMrpPWDWRE....DpDHFeVrMDMPGFKPEEIKVkvEDNNVLvIeG
               A   P++ R   + ++V++++ FKPEE+ VK+ D+ +++++G
Query         77  ARFGVPAEGR-TPPFPFGEPPWKVCNVVHSFKPEELMVTKDKG-YVEVSG 123

HMM            EHEREEEREDDKWWHERIYRHFMRFRrLPENVDPDqIkAsMSdNGVLTII
               +HE E++   + + ++ F   ++LP +VDP + AS+S++G+L I
Query         124 KHE---EKQQ----EGGIVSKNFTKKIQLPAEVDPTVTFASLSPEGLLI 166

HMM            TVPKpEP*
               ++P ++P
Query         167 EAPQVPP 173

```

DKF2phut1\_23g11

group: uterus derived

DKF2phut1\_23g11 encodes a novel 256 amino acid protein with similarity to S.pombe SPAC31G5.12c and S. cerevisiae Maf1p.

No informative BLAST results; No predictive prosite, pfam or SCOP motife.

The new protein can find application in studying the expression profile of uterus-specific genes.

similarity to SPAC31G5.12c and Maf1p

complete cDNA, complete cds, EST hits

Sequenced by EMBL

Locus: unknown

Insert length: 1674 bp

Poly A stretch at pos. 1664, polyadenylation signal at pos. 1644

```
1 GGGGGAGGCG GAGGTCGCTC GCTCGCTCGC TCGGCTCGCT GACTCGCCGG
51 AGCGCTCTGT GCGGTCGCGC GGCAGGTCGG TCGCGAGAGC GGGCTCTGTG
101 GAAGGGGGCG AGGCTATGTC GCGGTGGCAG CCCGATGGG CCGGCAGGGC
151 CGGGAGTAAC GGGACGTCGC CGCGGAGCTT CTTCCCCCGG ATACAGTGCG
201 GCCCGAGCGG AGGCCGCGGC GCGGCCCTCC GATCTTGAAG AGCCCCGCGT
251 GCGCGGAGCC CGCCCCCGCC TGCGCACCAG CACCGACGCG GAGCGACCAG
301 CCCAGCCAGA CCCGGCCCCG CGCGGCTGA TCTAACCCAG CCAGGCAGGC
351 AATACTAGCC CCTCTGGAGC ACGGAGCTCC TTCCCCAAG ACATGAAGCT
401 ATTGGAGAAC TCGAGCTTTG AAGCCATCAA CTCACAGCTG ACTGTGGAGA
451 CCGGAGATGC CCACATCATT GGCAGGATTG AGAGTACTC ATGTAAGATG
501 GCAGGAGAGC ACAAACACAT GTTCAAGCAG TTCTGCCAGG AGGGCCAGCC
551 CCACGTGCTG GAGGCACTTT CTCACCCCCA GACTTCAGGA CTGAGCCCCA
601 GCAGACTCAG CAAAGGCCAA GCGGTGAGG AGGAGGGCCC CCTCAGTGAC
651 AAGTGCAGCC GCAAGACCCT CTTCTACCTG ATTGCCACGC TCAATGAGTC
701 CTTCAGGCCCT GACTATGACT TCAGCACAGC CCGCAGCCAT GAGTTCAGCC
751 GGGAGGCCAG CCTTAGCTGG GTGGTGAATG CAGTCAACTG CAGTCTGTTC
801 TCAGCTGTGC GGGAGGACTT CAAGGATCTG AAACCACAGC TGTGGAACGC
851 GGTGGACGAG GAGATCTGCC TGGCTGAATG TGACATCTAC AGCTATAACC
901 CAGACTTGGA CTCAGATCCC TTCGGGGAGG ATGGTAGCCT CTGGTCCCTC
951 AACTACTTCT TCTACAACAA GCGGCTCAAG CGAATCGTCT TCTTTAGCTG
1001 CCGTTCCATC AGTGGCTCCA CCTACACACC CTCAGAGGCA GGCAACGAGC
1051 TGGACATGGA GCTGGGGGAG GAGGAGGTGG AGGAAGAAAG CAGAAGCAGG
1101 GGCAGTGGGG CCGAGGAGAC CAGCACCATG GAGGAGGACA GGGTCCAGT
1151 GATCTGTATT TGATGAGGAG GAGCCGAGGC CCCAGCTTCA TCCAGCTTCA
1201 ACCAATGCCT GGACCTGTCC ACCTGAGAGG CCCCTGGGGC CTCCTCAGCT
1251 GCTGGCCAGA CCTTGGCGCT GCCACAGTCC TGGCACTGCC CAAGGCCATA
1301 CCTGCTAGC CTTTGGCTC CATCTGTGG ATGCCCACTC ACCCTCAGA
1351 CTCTCTGCTG CCAATGCTGTG GCGGACTTGG TCAGCAGGGG GCCTGGTGGG
1401 AGGAGGAGCT GCCCTGCCCA AATGAATGCG CACAGCAGGG ACAGCTGGAC
1451 CGCAGAGTTT ATTTTGTAT TTCTACTGGG CCTGCACACT CCAGCCCCAA
1501 GGTCTGTGG CCGGAGGCCC CACGAGCAGG CCCAGCAGT CACCGGCTCT
1551 GGTCTTGGG CCGCCCCGGT GCCCACCTGT ACCCCACCT CGCCCATTTG
1601 GCCGCTGCA CTGAGTGCA CTTTGTGCTG GCTCGTTTCT TTCCAATAAA
1651 AGTTTCTGTG ACTTAAAAAA AAAA
```

## BLAST Results

No BLAST result

## Medline entries

No Medline entry

## Peptide information for frame 3

ORF from 393 bp to 1160 bp; peptide length: 256  
Category: similarity to known protein



BLASTP hits

Entry AF098499.2 from database TREMBL:  
gene: "C43H8.2"; *Caenorhabditis elegans* cosmid C43H8.  
Score = 263. P = 9.2e-23. identities = 78/252, positives = 118/252

No Alert BLASTP hits found

Pedant information for DKFZphutel 23q11, frame 3

## Report for DKFZphutel\_23g11.3

```
[LENGTH]      256
[MW]           28869.95
[pI]           4.51
[HOMOL]        TREMBL:SPAC31G5_12 gene: "SPAC31G5.12c"; product: "hypothetical protein";
S.pombe chromosome I cosmid c31G5. 4e-23
[FUNCAT]       06.04 protein targeting, sorting and translocation [S. cerevisiae, YDR005c]
6e-13
[PROSITE]      MYRISTYL           3
[PROSITE]      CK2_PHOSPHO_SITE      5
[PROSITE]      PKC_PHOSPHO_SITE      6
[PROSITE]      ASN_GLYCOSYLATION     3
[KW]           All_Alpha
[KW]           LOW_COMPLEXITY        7.81 %
```

```
SEQ      EETSTMEEDRVPVICI
SEG      XX.....
PRD      CCCCCCCCCCeeccC
```

## Prosites for DKFZphut1\_23g11.3

PS00001	6->10	ASN_GLYCOSYLATION	PDOC00001
PS00001	101->105	ASN_GLYCOSYLATION	PDOC00001
PS00001	132->136	ASN_GLYCOSYLATION	PDOC00001
PS00005	33->36	PKC_PHOSPHO_SITE	PDOC00005
PS00005	85->88	PKC_PHOSPHO_SITE	PDOC00005
PS00005	89->92	PKC_PHOSPHO_SITE	PDOC00005
PS00005	103->106	PKC_PHOSPHO_SITE	PDOC00005
PS00005	112->115	PKC_PHOSPHO_SITE	PDOC00005
PS00005	202->205	PKC_PHOSPHO_SITE	PDOC00005
PS00006	7->11	CK2_PHOSPHO_SITE	PDOC00006
PS00006	99->103	CK2_PHOSPHO_SITE	PDOC00006
PS00006	212->216	CK2_PHOSPHO_SITE	PDOC00006
PS00006	238->242	CK2_PHOSPHO_SITE	PDOC00006
PS00006	244->248	CK2_PHOSPHO_SITE	PDOC00006
PS00008	66->72	MYRISTYL	PDOC00008
PS00008	181->187	MYRISTYL	PDOC00008
PS00008	239->245	MYRISTYL	PDOC00008

(No Pfam data available for DKFZphut1\_23g11.3)

DKFZphut1\_24c19

group: transmembrane protein

DKFZphut1\_24c19 encodes a novel 195 amino acid protein without similarity to known proteins.

The novel protein contains 1 transmembrane region.  
No informative BLAST results; No predictive prosite, pfam or SCOP motife.

The new protein can find application in studying the expression profile of uterus-specific genes and as a new marker for uterine cells.

unknown

membrane regions: 1

Summary DKFZphut1\_24c19 encodes a novel 195 amino acid protein, with  
no similarity to known proteins.

unknown

complete cDNA, complete cds, EST hits  
TRANSMEMBRANE 1

Sequenced by Qiagen

Locus: unknown

Insert length: 769 bp

Poly A stretch at pos. 746, polyadenylation signal at pos. 735

```
1 ACGAGTCAGC CAAAGATGGC TGC GCCCAGG TAATTGAGC AAAGGCCACA
51 GTGAACTCCG GCGTGGCTGA GGAAGACCGG AGGAGGCACC CACAGGCTGC
101 TGGGAGGAGA GCATAAGGCT CAAAATGGAA AATCATAAAT CCAATAATAA
151 GGAAACATA ACAATTGTTG ATATATCCAG AAAAATTAAC CAGCTTCCAG
201 AAGCAGAAAG GAATCTACTT GAAAATGGAT CGGTTTATGT TGGATTAAAT
251 GCTGCTCTTT GTGGCCTCAT AGCAACAGT CTTTTCGAC GCATCTTGAA
301 TGTGACAAAG GCTCGCATAG CTGCTGGCTT ACCAATGGCA GGGATACCTT
351 TTCTTACAAC AGACTTAACT TACAGATGTT TTGTAAGTTT TCCTTTGAAT
401 ACAGGTGATT TGGATTGTGA AACCTGTACC ATAACACGGA GTGGACTGAC
451 TGGTCTTGTT ATTGGTGGTC TATACCTGT TTTCTTGGCT ATACCTGTAA
501 ATGGTGGTCT AGCAGCCAGG TATCAATCAG CTCTGTTACC ACACAAAGGG
551 AACATCTTAA GTTACTGGAT TAGAACTTCT AAGCCTGTCT TTAGAAAGAT
601 GTTATTTTCT ATTTTGCTCC AGACTATGTT TTCAGCATAC CTGGGTCTG
651 AACAAATATA ACTACTTATA AAGGCCCTTC AGTTATCTGA ACCTGGCAAA
701 GAAATTCACG GATTTTAAAC AAATATGTAA ACAAATAATA AATGGTAAAA
751 ACAAAAAAAA AAAAAAAA
```

## BLAST Results

No BLAST result

## Medline entries

No Medline entry

## Peptide information for frame 2

ORF from 125 bp to 709 bp; peptide length: 195  
Category: putative protein

```
1 MENHKSNNKE NITIVDISRK INQLPEAERN LLENGSVYVG LNAALCGLIA
51 NSLFRRILNV TKARIAAGLP MAGIPFLTDD LTYRCFVSFP LNTGDLDCET
101 CTITRSGLTG LVIGGLYPVF LAIPVNGGLA ARYQSALLPH KGNILSYWIR
151 TSKPVFRKML FPILLQTMFS AYLGSEQYKL LIKALQLSEP GKEIH
```

## BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphut1\_24c19, frame 2

No Alert BLASTP hits found

Pedant information for DKFZphut1\_24c19, frame 2

Report for DKFZphut1\_24c19.2

```

[LENGTH]      195
[MW]           21527.45
[pI]           9.36
[PROSITE]      MYRISTYL      6
[PROSITE]      CK2_PHOSPHO_SITE      1
[PROSITE]      PKC_PHOSPHO_SITE      3
[PROSITE]      ASN_GLYCOSYLATION      3
[KW]           TRANSMEMBRANE 1

SEQ  MENHKSNNKENITIVDISRKINQLPEAERNLLENGSVYVGLNAALCGLIANSIFRRILNV
PRD  ccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc
MEM  .....

SEQ  TKARIAAGLPMAGIPFLTTLTYRCFVSFPLNTGDLDCETCTITRSGTLGLVIGGLYPVF
PRD  hhhhhhcccccccccccccccccccccccccccccccccccccccccccccccccccccccc
MEM  .....MMMMMMMMMMMMMM

SEQ  LAIPVNGGLAARYQSALLPHKGNILSYWIRTSKPVFRKMLFPIQLTMSAYLGSEQYKL
PRD  eeeeecccccccccccccccccccccccccccccccccccccccccccccccccccccccc
MEM  MMM.....

SEQ  LIKALQLSEPGKEIH
PRD  hhhhhhcccccccc
MEM  .....

```

Prosite for DKFZphut1\_24c19.2

```

PS00001      11->15  ASN_GLYCOSYLATION      PDOC00001
PS00001      34->38  ASN_GLYCOSYLATION      PDOC00001
PS00001      59->63  ASN_GLYCOSYLATION      PDOC00001
PS00005      18->21  PKC_PHOSPHO_SITE      PDOC00005
PS00005      82->85  PKC_PHOSPHO_SITE      PDOC00005
PS00005     151->154 PKC_PHOSPHO_SITE      PDOC00005
PS00006      13->17  CK2_PHOSPHO_SITE      PDOC00006
PS00008      40->46  MYRISTYL      PDOC00008
PS00008      47->53  MYRISTYL      PDOC00008
PS00008      68->74  MYRISTYL      PDOC00008
PS00008     110->116 MYRISTYL      PDOC00008
PS00008     127->133 MYRISTYL      PDOC00008
PS00008     142->148 MYRISTYL      PDOC00008

```

(No Pfam data available for DKFZphut1\_24c19.2)

DKFZphute1\_24e11

-----

group: intracellular transport and trafficking

DKFZphute1\_24e11 encodes a novel 226 amino acid protein, with similarity to human/mouse golgi 4-transmembrane spanning transporter MTP. MTP may function in the transport of nucleosides and/or nucleoside derivatives between the cytosol and the lumen of an intracellular membrane-bound compartment. Thus, the novel protein also seems to be involved in nucleotide sugar transport.

The new protein can find application in modulating the transport of nucleosides and/or nucleoside derivatives between the cytosol and the lumen of an intracellular membrane-bound compartments.

similarity to 4-TRANSMEMBRANE SPANNING TRANSPORTER MTP

complete cDNA, complete cds, EST hits

potential start at 184,

TRANSMEMBRANE 4

function in the transport of nucleosides and/or nucleoside derivatives between the cytosol and the lumen of an intracellular membrane-bound compartment?

Sequenced by Qiagen

Locus: /map="8"

Insert length: 2005 bp

Poly A stretch at pos. 1988, polyadenylation signal at pos. 1963

```

1  ACGCGTCCGG CAGAAGCTCG GAGCTCTCGG GGTATCGAGG AGGCAGGCC
51  GCGGGCGCAC GGGCGAGCGG GCCGGGAGCC GGAGCGGCGG AGGAGCCGGC
101 AGCAGCGCGC CGGCGGGCTC CAGGCGAGGC GGTGACGCTC CTTGAAACT
151 TCGCGCGCGC CTCGCGCCAC TCGCGCCGGA GCGATGAAGA TGCTCGCGCC
201 CTGGACGCGG TTCTACTCCA ACAGCTGCTG CTTGTGCTGC CATGTCCGCA
251 CCGGCACCAT CCGTCTCGGC GTCTGGTATC TGATCATCAA TGCTGTGGTA
301 CTGTTGATTT TATTGAGTGC CCGGCTGATC CCGATCAGT  ATAACCTTTC
351 AAGTTCTGAA CTGGGAGGTG ACTTTGAGTT CATGGATGAT GCCAACATGT
401 GCATTGCCAT TGGGATTTCT CTTCTCATGA TCCTGATATG TGCTATGGCT
451 ACTTACGGAG CGTACAAGCA ACGCGCAGCC TGGATCATCC CATCTCTCTG
501 TTACACAGAT TTTGACTTTG CCTGAACAT GTTGGTTGCA ATCACTGTGC
551 TTATTTATCC AAATCCATT CAGGAATACA TACGGCAACT GCCTCCTAAT
601 TTTCCCTACA GAGATGATGT CATGTCAGTG AATCCTACCT GTTTGGTCTT
651 TATTATTCTT CTGTTTATTA GCATTATCTT GACTTTTAAG GGTACTTGA
701 TTAGCTGTGT TTGGAACATG TACCGATACA TCAATGGTAG GAACTCCTCT
751 GATGTCCTGG TTTATGTTAC CAGCAATGAC ACTACGGTGC TGCTACCCCC
801 GTATGATGAT GCCACTGTGA ATGGTGCTGC CAAGGAGCCA CCGCCACCTT
851 ACGTGTCTGC CTAAGCCTTC AAGTGGGCGG AGCTGAGGGC AGCAGCTTGA
901 CTTTGCAGAC ATCTGAGCAA TAGTTCGTGT ATTTCACTTT TGCCATGAGC
951 CTCTCTGAGC TTGTTTGTG CTGAAATGCT ACTTTTAAA ATTTAGATGT
1001 TAGATTGAAA ACTGTAGTTT TCAACATATG CTTTGCTAGA ACACGTGTGAT
1051 AGATTAACTG TAGAATCTT CCGTACGATG TGGGATATA ACGGGCTTCA
1101 CTAACCTTCC CTAGGCATTG AAATCTCCC CAAATCTGAT GGACCTAGAA
1151 GTCTGCTTTT GTACCTGCTG GGGCCCAAAG TTGGGCATTT TTCTCTCTGT
1201 TCCCTCTCTT TTGAAAATGT AAAATAAAAC CAAAAATAGA CAACCTTTTC
1251 TTCAGCCATT CCAGCATAGA GAACAAACC TTATGGAAAC AGGAATGTCA
1301 ATTTGTGAAT CATTGTTCTA ATTAGGTAAG TAGAAGTCTT TATGTATGTG
1351 TTACAAGAAAT TTCCCCACAC ACATCCTTTA TGAATGAAGT TCAATGACAG
1401 TTTGTGTTTG GTGGTAAAGG ATTTCTCCA TGGCTGAAT TAAGACCAT
1451 AGAAAGCACC AGGCGTGGG AGCAGTGACC ATCTACTGAC TGTTCTTGTG
1501 GATCTTGTGT CCAGGGACAT GGGGTGACAT GCCTCGTATG TGTAGAGGG
1551 TGGAAATGGAT GTGTTTGGCG CTGCATGGGA TCTGGTGCCC CTCTTCTCCT
1601 GGATTACATC CCCACCCAG GGGCCGCTTT TACTAAGTGT TCTGCCCTAG
1651 ATTTGTTCAA GGAGGTCATC CAACTGACTT TATCAAGTGG AATTGGGATA
1701 TATTTGATAT ACTTCTGCTT AACAACTGG AAAAGGGTTT TCTTTTCCCT
1751 GCAAGCTACA TCCTACTGCT TTGAACCTCC AAGTATGTCT AGTCACCTTT
1801 TAAAATGTAA ACATTTTCAG AAAAATGAGG ATTGCCTTCC TTGTATGCGC
1851 TTTTACCTT GACTACCTGA ATTGCAAGG ATTTTATAT ATTCATATGT
1901 TACAAAGTCA GCAACTCTCC TGTGGTTTCA TTATTGAATG TGCTGTAAAT
1951 TAAGTCGTTT GCAATTAAAA CAAGGTTTGC CCACATCCAA AAAAAAAAAA
2001 AAAAA

```

## BLAST Results

-----

Entry HS012351 from database EMBL:

human STS SHGC-31823.  
Score = 1629, P = 3.1e-67, identities = 343/354

## Medline entries

96199248:  
Identification of a novel membrane transporter  
associated with intracellular membranes by  
phenotypic complementation in the yeast  
Saccharomyces cerevisiae.

## Peptide information for frame 1

ORF from 184 bp to 861 bp; peptide length: 226  
Category: strong similarity to known protein

1 MKMVAPWTRF YSNSCCLCCH VRTGTILLGV WYLIINAVVL LILLSALADP  
51 DQYNFSSSEL GGDFFEMDDA NMCIAIAISL LMILICAMAT YGAYKQRAAW  
101 IIPFFCYQIF DFALNMLVAI TVLIYPNSIQ EYIRQLPPNF PYRDDVMSVN  
151 PTCLVLIILL FISIILTFKG YLISCVWNCY RYINGRNSSD VLVYVTSNDT  
201 TVLPPYDDA TVNGAAKEPP PPYVSA

## BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphut1\_24e11, frame 1

SWISSPROT:MTRP\_HUMAN GOLGI 4-TRANSMEMBRANE SPANNING TRANSPORTER MTP  
(KIAA0108)., N = 1, Score = 551, P = 2.9e-53

SWISSPROT:MTRP\_MOUSE GOLGI 4-TRANSMEMBRANE SPANNING TRANSPORTER MTP., N  
= 1, Score = 539, P = 5.3e-52

TREMBL:HS304981\_1 product: "E3 protein"; Human retinoic acid-inducible  
E3 protein mRNA, complete cds., N = 1, Score = 127, P = 3.4e-06

>SWISSPROT:MTRP\_HUMAN GOLGI 4-TRANSMEMBRANE SPANNING TRANSPORTER MTP  
(KIAA0108).  
Length = 233

## HSPs:

Score = 551 (82.7 bits), Expect = 2.9e-53, P = 2.9e-53  
Identities = 102/221 (46%), Positives = 148/221 (66%)

Query: 9 RFYSNSCCLCCHVRTGTILLGVWYLIINAVVLLILLSALADPDQY---NFSSELGGDF- 64  
RFYS CC CCHVRTGTI+LG WY+++N ++ ++L + P+ N +G +  
Sbjct: 13 RFYSTRCCGCCHVRTGTIILGTWYVMVNNLLMAILLTVEVTHPNSMPAVNIQYEVIGNYYS 72  
Query: 65 -EFMDANMCIAIAISLLMILICAMATYGAYKQRAAWIIPFFCYQIFDFALNMLVAITVL 123  
E M D N C+ A+S+LM +I +M YGA + W+IPFFCY++FDF L+ LVAT+ L  
Sbjct: 73 SERMAD-NACVLFVAVSVLMFISSMLVYGAISYQVQWLIIPFFCYRLDFVLSCLVAISSL 131  
Query: 124 IYPNSIQEYIRQLPPNFYRDDVMSVNPCLVLIILLFISIILTFKGYLISCVWNCYRYI 183  
Y I+EY+ QLP +FPY+DD++++ +CL+ I+L+F ++ + FK YLI+CVWNCY+YI  
Sbjct: 132 TYLPRIKEYLDQLP-DFPYKDDLLALDSSCLLFIVLVFFALFIIFKAYLINCWNCYKYI 190  
Query: 184 NGRNSSDVLVYVTSN-DTTVLLPPYDDATVNGAAKEPPPPYVSA 226  
N RN ++ VY +LP Y+ A V KEPPPPY+ A  
Sbjct: 191 NNRNVPEIAVYPAFEAPPQYVLPTYEMA-VKMPEKEPPPPYLPA 233

## Pedant information for DKFZphut1\_24e11, frame 1

## Report for DKFZphut1\_24e11.1

[LENGTH] 226  
[MW] 25419.11

[pI] 4.65  
 [HOMOL] SWISSPROT:MTRP\_HUMAN GOLGI 4-TRANSMEMBRANE SPANNING TRANSPORTER MTP (KIAA0108).  
 5e-40  
 [PROSITE] CK2\_PHOSPHO\_SITE 3  
 [PROSITE] TYR\_PHOSPHO\_SITE 1  
 [PROSITE] PKC\_PHOSPHO\_SITE 1  
 [PROSITE] ASN\_GLYCOSYLATION 3  
 [KW] SIGNAL PEPTIDE 49  
 [KW] TRANSMEMBRANE 2  
 [KW] LOW\_COMPLEXITY 20.80 %

SEQ MKMVAPWTRFYSNSCLCCHVRTGTILLGVWYLIINAVVLLILLSALADPDQYNFSSSEL  
 SEG .....XXXXXXXXXXXXXXXXXXXXX.....  
 PRD ccc  
 MEM .....

SEQ GGDFEFMDANMCIAIAISLLMILICAMATYGAYKQRAAWIIPFCYQIFDFALNMLVAI  
 SEG .....XXXXXXXXXXXXXXXXXXXXX.....  
 PRD ccc  
 MEM .....MM

SEQ TVLIYPNSIQEYIRQLPPNFYRDDVMSVNPTCLVLIILLFISIIILTFKGYLISCVWNCY  
 SEG .....XXXXXXXXXXXXXXXXXXXXX.....  
 PRD hhhcc  
 MEM MMMMM.....MM...

SEQ RYINGRNSSDVLVYVTSNDTTVLLPPYDDATVNGAAKEPPPPYVSA  
 SEG .....  
 PRD ecc  
 MEM .....

#### Prosites for DKFZphute1\_24e11.1

PS00001	54->58	ASN_GLYCOSYLATION	PDOC00001
PS00001	187->191	ASN_GLYCOSYLATION	PDOC00001
PS00001	198->202	ASN_GLYCOSYLATION	PDOC00001
PS00005	167->170	PKC_PHOSPHO_SITE	PDOC00005
PS00006	56->60	CK2_PHOSPHO_SITE	PDOC00006
PS00006	128->132	CK2_PHOSPHO_SITE	PDOC00006
PS00006	196->200	CK2_PHOSPHO_SITE	PDOC00006
PS00007	186->195	TYR_PHOSPHO_SITE	PDOC00007

(No Pfam data available for DKFZphute1\_24e11.1)

DKF2phutel\_24j6

group: cell structure and motility

DKF2phutes1\_24j6 encodes a novel 571 amino acid protein with strong similarity to rat cell adhesion regulator (CARL).

The novel protein is very similar to Carl and thus seems to be involved in regulation cell-cell adhesion. It contains a RGD cell attachment site.

The new protein can find application in modulation of cell-cell-adhesion.

strong similarity to rat CARL A.thaliana T19C21.5

complete cDNA, complete cds, EST hits  
potential frame shift at Bp 1241 according to CARL  
but frame shift might be in CARL sequence!  
ESTs T73366 AA362984 confirm this sequence

Sequenced by Qiagen

Locus: /map="939.9 cR from top of Chr2 linkage group"

Insert length: 3333 bp

Poly A stretch at pos. 3316, no polyadenylation signal found

```
1  ACGCGTCCGA GCTGGCTCAG GCGCTCCGCT AGGCTCGGAC GACCTGCTGA
51  GCCTCCCAAA CCGCTTCCAT AAGGCTTTGC CTTTCCAACT TCAGCTACAG
101 TGTTAGCTAA GTTTGGAAAG AAGGAAAAAA GAAATCCCTT GGGCCCCCTT
151 TCTTTTGTTC TTTGCCAAAG TCGTCGTTGT AGTCTTTTGT CCCAAGGCTG
201 TTGTGTTTTT AGAGGTGCTA TCTCCAGTTC CTGCACTCC TGTTAAACAAG
251 CACCTCAGCG AGAGCAGCAG CAGCGATAGC AGCCGAGAA GAGCCAGCGG
301 GGTCCGCTAG TGTCATGACC AGGCGGGAG ATCACAACCG CCAGAGAGGA
351 TGCTGTGGAT CCTTGGCCGA CTACCTGACC TCTGCAAAAT TCCTTCTCTA
401 CCTTGGTCAT TCTCTCTCTA CTGGGGGAGA TCGGATGTGG CACTTTGCGG
451 TGTCGTGTTT TCTGGTAGAG CTCTATGAA ACAGCCTCCT TTTGACAGCA
501 GTCTACGGGC TGGTGGTGGC AGGGTCTGTT CTGGTCTGG GAGCCATCAT
551 CGGTGACTGG GTGGACAAGA ATGCTAGACT TAAAGTGGCC CAGACCTCGC
601 TGGTGGTACA GAATGTTTCA GTCATCCTGT GTGGAATCAT CCTGATGATG
651 GTTTTCTTAC ATAAACATGA GCTTCTGACC ATGTACCATG GATGGGTCTT
701 CACTTCTTGC TATATCTTGA TCATCACTAT TGCAAAATAT GCAAAATTGG
751 CCAGTACTGC TACTGCAATC ACAATCCAAA GGGATTGGAT TGTGTTGTT
801 CGAGGAGAAAG ACAGAAGCAA ACTAGCAAA ATGAATGCCA CAATACGAAG
851 GATTGACCAG TTAACCAACA TCTTAGCCCC CATGGCTGTT GGCCAGATTA
901 TGACATTTGG CTCCCCAGTC ATCGGCTGTG GCTTTATTTT GGGATGGAAC
951 TTGGTATCCA TGTGCGTGGG GTACGTCTCT CTCTGGAAGG TTTACCAGAA
1001 AACCCACGCT CTAGCTGTGA AAGCTGGTCT TAAAGAAGAG GAAACTGAAT
1051 TGAAACAGCT GAATTTACAC AAAGATACTG AGCCAAACCC CCTGGAGGGA
1101 ACTCATCTAA TGGGTGTGAA AGACTCTAAC ATCCATGAGC TTGAACATGA
1151 GCAAGAGCCT ACTTGTGCCT CCCAGATGGC TGAGCCCTTC CGTACCTTCC
1201 GAGATGGATG GGTCTCTTAC TACAACAGC CTGTGTTTCT GGCTGGCATG
1251 GGTCTTGTCT TCCTTTATAT GACTGTCTGT GGCTTTGACT GCATCACCAC
1301 AGGGTACGCC TACACTCAGG GACTGAGTGG TTCCATCTCT AGTATTTTGA
1351 TGGGAGCATC AGCTATAACT GGAATAATGG GAACGTAGC TTTTACTTGG
1401 CTACGTCGAA AATGTGGTTT GGTTCGGACA GGTCTGATCT CAGGATTGGC
1451 ACAGCTTTCC TGTTTGATCT TGTGTGTGAT CTCTGTATTC ATGCCTGGAA
1501 GCCCCTTGA CTGTCCGTT TCTCCTTTTG AAGATATCCG ATCAAGGTTT
1551 ATTCAAGGAG AGTCAATTAC ACCTACCAAG ATACCTGAAA TTACAACATGA
1601 AATATACATG TCTAATGGGT CTAATTCTGC TAATATTGTC CCGGAGACAA
1651 GTCCTGAATC TGTGCCATA ATCTCTGTCA GTCTGCTGTT TGCAGGCGTC
1701 ATTGCTGCTA GAATCGGTCT TTGGTCCTTT GATTTAACTG TGACACAGTT
1751 GCTGCAAGAA AATGTAATTG AATCTGAAAG AGGCATTATA AATGGTGTAC
1801 AGAAGTCCAT GAACTATCTT CTGTATCTTC TGCATTTCAT CATGGTCATC
1851 CTGGCTCCAA ATCCTGAAGC TTTTGGCTTG CTGATTTGA TTTCACTCTC
1901 CTTTGTGGCA ATGGGCCACA TTATGTATTT CCGATTTGCC CAAATATCTC
1951 TGGGAAACAA GCTCTTTGCT TGCGGTCTCT ATGCAAAAGA AGTTAGGAAG
2001 GAAATCAAG CAAATACATC TGTTGTTTGA GACAGTTTAA CTGTTGCTAT
2051 CCTGTACTTA GATTATATAG AGCAGATGTG CTTATTTTGT ACTGCAGAA
2101 TCCAATAAAT GGCTGGGTGT TTTGCTCTGT TTTTACCACA GCTGTGCCCT
2151 GAGAACTAAA AGCTGTTTAG GAAACCTAAG TCAGCAGAAA TAACTGATT
2201 AATTTCCCTT ATGTTGAGGC ATGGAAGAAA AATGGGAAA GAAAACTCA
2251 GTTTAAATAC GGAGACTATA ATGATAACAC TGAATTCCCC TATTCTCAT
2301 GAGTAGATAC AATCTTACGT AAAAGAGTGG TTAGTCACGT GAATTCAGTT
2351 ATCATTGAC AGATTCTTAT CTGTACTAGA ATTGAGATAT GTCAGTTTTC
2401 TGCAAAACTC ACTCTTGTTC AAGACTAGCT AATTTATTTT TTTGCATCTT
2451 AGTTATTTT AAAACAAAT TCTTCAAGTA TGAAGACTAA ATTTTGATA
2501 CTAATATTAT CCTATTGAT CCTATTGATC TTAAGGTATT TACATGTATG
```



```

2551 TGGAAAAACA AAACACTTAA CTAGAATTCT CTAATAAGGT TTATGGTTTA
2601 GCTTAAAGAG CACCTTTGTA TTTTATTAT CAGATGGGGC AACATATTGT
2651 ATGAAGCATA TGTAAGCACT CACAGCATGG TTATCATGTA AGCTGCAGGT
2701 AGAAGCAAAG CTGTAAAGTA GATTATACAC ACAATGACTG CATACAGACT
2751 TCAATATATGT CAATAGTTTG GTCATAGAAC CTAGAAGCCA AAAGCCACAC
2801 AGAAGGGCAA GAATCCCAAT TTAACCTCAT TTATCATCAT TAGTGATCTG
2851 TGTGTAGAAA CATGAGGGTG TAAGCCTTCA GCCTGGCAAG TTACATGTAG
2901 AAAGCCCAACA CTTGTGAAGG TTTTGTTTA CAAATCACTT GATTAAACAC
2951 ACTCAGGTAG AATATTTTAA TTTTACTGT TTTATACCCA GAAGTTATTT
3001 CTACATTGTT CTACAGCAAG AATATTCATA AAAGTATCCC TTCAAATGC
3051 CTTTGAGAAG AATAGAAGAA AAAAAGTTTG TATATATTTT AAAAAATTGT
3101 TTTAAAAGTC AGTTTGCAAC ATGCTGTGAC CAAGATGGTA CTTGCGCTTA
3151 ACCGTTTATA TGCACCTTCA TGGAGACTGC AATACGTTGC TATGAGCACT
3201 TCTTTTATCC TTGGAGTTTA ATCCTTTGCT TCATCTTTCT ACAGTATGAC
3251 ATAATGATT GCTATGTTGT AAAATCTTTG TAAAAAATTT CTATATAAAA
3301 ATATTTTGAA AATCTTAAAA AAAAAAAAAA AAA

```

## BLAST Results

Entry HS389210 from database EMBL:  
human STS SHGC-10164.  
Score = 1592, P = 1.5e-64, identities = 346/364

Entry HS933343 from database EMBL:  
human STS WI-16551.  
Score = 1193, P = 5.7e-46, identities = 241/244

## Medline entries

No Medline entry

## Peptide information for frame 3

ORF from 315 bp to 2027 bp; peptide length: 571  
Category: strong similarity to known protein

```

1 MTRAGDHNRO RGCCGSLADY LTSAKFLLYL GHSLSWTGDR MWHFAVSVFL
51 VELYGNLLLL TAVYGLVVAG SVLVLGAIIG DWVDKNARLK VAQTSLVVQN
101 VSVILCGIIL MMVFLHKKHEL LTMVHGWLVT SCYILITIA NIANLASTAT
151 AITIQRDWIV VVAGEDRSKL ANMNATIRRI DQLTNIALPM AVGQIMTFGS
201 PVIGCGFISG WNLVSMCWEY VLLWKVYQRT PALAVKAGLK EEETELKQLN
251 LHKDTEPKPL EGTLMGVKD SNIHELEHEQ EPTCASQMAE PFRTFRDQWV
301 SYYNQPVFLA GMGLAFLYMT VLGFDCTTG YAYTQGLSGS ILSILMGASA
351 ITGIMGTVAE TWLRRKGLV RTGLISGLAQ LSCLILCVIS VFMPGSLDL
401 SVSPFEDIRS RFIQGESITP TKIPEITTEI YMSNGSNSAN IVPETSPESV
451 PIISVSLLEA GVIAARIGLW SFDLTVTQLL QENVIESERG IINGVQNSMN
501 YLLDLLHFIM VILAPNPEAF GLLVLISVSF VAMGHIMYFR FAQNTLGNKL
551 FACGPDAKEV RKENQANTSV V

```

## BLASTP hits

No BLASTP hits available

## Alert BLASTP hits for DKF2phut1\_24j6, frame 3

TREMBLNEW:U76714\_1 gene: "CAR1"; product: "cell adhesion regulator";  
Rattus norvegicus cell adhesion regulator (CAR1) mRNA, complete cds., N  
= 1, Score = 1472, P = 7.2e-151

TREMBL:AC004683\_5 gene: "T19C21.5"; Arabidopsis thaliana chromosome II  
BAC T19C21 genomic sequence, complete sequence., N = 2, Score = 437, P  
= 2.8e-60

TREMBL:AF039046\_2 gene: "R09B5.4"; Caenorhabditis elegans cosmid  
R09B5., N = 2, Score = 323, P = 1.5e-43

>TREMBLNEW:U76714\_1 gene: "CAR1"; product: "cell adhesion regulator";  
Rattus norvegicus cell adhesion regulator (CAR1) mRNA, complete cds.  
Length = 405

Score = 1472 (220.9 bits), Expect = 7.2e-151, P = 7.2e-151  
Identities = 288/319 (90%), Positives = 297/319 (93%)

Query:	1	MTRAGDHNRRQGCCGLADYLTSAKFLLYLGHSLSTWGDMMWHFAVSVFLVELYGNLSLL	60
		MT++ D Q CGCCSLA+YLTSAKFLLYLGHSLSTWGDMMWHFAVSVFLVELYGNLSLL	
Sbjct:	1	MTKSRDDTHQEGGCCSLANYLTSAKFLLYLGHSLSTWGDMMWHFAVSVFLVELYGNLSLL	60
Query:	61	TAVYGLVVGASVVLVLAGIIGDWVDKNARLKAQTSLVVQNVSVILCGIILMMVFLHKHEL	120
		TAVYGLVVGASVVLVLAGIIGDWVDKNARLKAQTSLVVQNVSVILCGIILMMVFLHK+EL	
Sbjct:	61	TAVYGLVVGASVVLVLAGIIGDWVDKNARLKAQTSLVVQNVSVILCGIILMMVFLHKNEL	120
Query:	121	LTMHYGHWLTSYILIITIANIANLASTATAITIQRDWIVVVGEDRSKLNMNATIRRI	180
		L MYHGWLVT CYILIITIANIANLASTATAITIQRDWIVVVGAE+RS+LA+MNATIRRI	
Sbjct:	121	LNMHYGHWLTVCYILIITIANIANLASTATAITIQRDWIVVVGAGNRSLADMNATIRRI	180
Query:	181	DQLTNILAPMAVGQIMTFGSPVIGCGFISGWNLSVMCEYVLLWVKVQKTPALAVKAGLK	240
		DQLTNILAPMAVGQIMTFGSPVIGCGFISGWNLSVMCEY LLWVKVQKTPALAVKA LK	
Sbjct:	181	DQLTNILAPMAVGQIMTFGSPVIGCGFISGWNLSVMCEYFLLWVKVQKTPALAVKAALK	240
Query:	241	EEETELKQLNLHKDTEPKPLEGTHLMCKVSDSNIHELEHEQEPTCASQMAEPFRTFRDGV	300
		EE+ELKQL KDTEPKPLEGTHLM KDSNI ELE EQEPTCASQ+AEPTFRDGV	
Sbjct:	241	VEESLKLQTSPKDTEPKPLEGTHLMGEKSDSNIRESLEPTCASQIAEPFRTFRDGV	300
Query:	301	SYYNQPVFLAGMGLAF-LY 318	
		SYYNQPVFL G FLY	
Sbjct:	301	SYYNQPVFLGWHGPGGFPLY 319	

Pedant information for DKF2phutel 24j6, frame 3

## Report for DKFZphute1 2416.3

```
[LENGTH]          571
[MW]               62542.72
[pI]              6.08
[HOMOL]           TREMBL:U76714.1 gene: "CAR1"; product: "cell adhesion regulator"; Rattus
norvegicus cell adhesion regulator (CAR1) mRNA, complete cds. 1e-141
[BLOCKS]          BL00341D
[PROSITE]         MYRISTYL      15
[PROSITE]         MITOCH_CARRIER 1
[PROSITE]         CK2_PHOSPHO_SITE      6
[PROSITE]         PROKAR_LIPOPROTEIN    1
[PROSITE]         PKC_PHOSPHO_SITE      4
[PROSITE]         ASN_GLYCOSYLATION      4
[PFAM]            Laminin B (Domain IV)
[KW]              TRANSMEMBRANE 4
[KW]              LOW COMPLEXITY      8.76 %
```

```

SEQ      MTRAGDHNRRQGCCGLADYLTSAKFLLYLGHSLSTWGDRMWHFAVSFVLVLYGNSLLL
SEG
PRD      cccccccccccccccccchhhhhhhhhheeeccceeeccccchhhhhhhhhheeecccccce
MEM      .MMMMMMMMMMMMMM

SEQ      TAVYGLVVGASVLVLGAIIGDWVDKNARLKVAQTSLVVQNVSVILCGIILMMVFLHKHEL
SEG      .xxxxxxxxxxxxxxxxxxxx
PRD      ehhhhhhhccceeeccccccchhhhhhhhhhhheeeccchhhhhhhhhhhhhhhhhhhhh
MEM      MMMMMMMMMMMMMMMM.MMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMM

SEQ      LTMYHGCVLTSCYILITIIANIANLASTATAITIQRDWIVVAGEDRSKLANMNATIRRI
SEG      .xxxxxxxxxxxxxxxxxxxxxxxxxxxx
PRD      hhccccchhhhhhhhhhhhhhhhhhhhhheeeccceeeccccchhhhhhhhhhhhhhh
MEM      MMMMMMMM

SEQ      DQLTNILAPMAVGQIMTFGSPVIGCGFISGWNLVSMCVEYVLLWKVYQKTPALAVKAGLK
SEG
PRD      hhhhhhhccceeeceeeeeeccceeeeeeccchhhhhhhhhhhhhhhhhhhccchhhhhhhhh
MEM

SEQ      EEETELKQLNLHKDTEPKPLEGTHLMGVKDSNIHELEHEQEPTCASQMAEPFRTFRDGVV
SEG
PRD      hhhhhhhhhhhccccccccccccceeeeeeccccccccccccccccccccccccccccce
MEM

SEQ      SYYNQPVFLAGMGLAFLYMTVLGFDICITTYAYTQGLSGSILSILMGASAITGIMGTAVF
SEG
PRD      eeccceeeccccchhhhhhhccccceeeeeeccccceeeeeeccccceeeehhhhhhh

```

```

MEM .....
SEQ  TWLRRKGLVVRTGLISGLAQLSCLILCVISVFMFGSPDLDSVSPFEDIRSRFIQGESITP
SEG  .....xxx
PRD  hhhhhccccccccchhhhhhhhhhhhhhhhhccccccccccccchhhhhcccccccc
MEM  .....

SEQ  TKIPEITTEIYMSNGSNSANIVPETSPEVPIISVSLFAGVIAARIGLWSFDLTVTQLL
SEG  xxxxxxxxxxxx.....
PRD  cccccceeeeeccccccccccccccccceeeeeehhhhhhhhhhhccccchhhhhhhhh
MEM  .....MMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMM.....

SEQ  QENVIESERGIINGVQNSMNYLLDLLHFIMVILAPNPEAFGLLVLSVSFVAMGHIMYFR
SEG  .....
PRD  hhhhhccccceeeeeccccchhhhhhhhhhhheeeeeccccccccceeeeeccccccccceeee
MEM  .....MMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMM.....

SEQ  FAQNTLGKGLFACGPDKEVRKENQANTSVV
SEG  .....
PRD  eccccccccceeeeeccccchhhhhhhhhcccccc
MEM  .....

```

## Prosites for DKF2phutet\_24j6.3

PS000001	100->104	ASN_GLYCOSYLATION	PDOC000001
PS000001	174->178	ASN_GLYCOSYLATION	PDOC000001
PS000001	434->438	ASN_GLYCOSYLATION	PDOC000001
PS000001	567->571	ASN_GLYCOSYLATION	PDOC000001
PS000005	23->26	PKC_PHOSPHO_SITE	PDOC000005
PS000005	176->179	PKC_PHOSPHO_SITE	PDOC000005
PS000005	294->297	PKC_PHOSPHO_SITE	PDOC000005
PS000005	487->490	PKC_PHOSPHO_SITE	PDOC000005
PS000006	16->20	CK2_PHOSPHO_SITE	PDOC000006
PS000006	36->40	CK2_PHOSPHO_SITE	PDOC000006
PS000006	294->298	CK2_PHOSPHO_SITE	PDOC000006
PS000006	396->400	CK2_PHOSPHO_SITE	PDOC000006
PS000006	403->407	CK2_PHOSPHO_SITE	PDOC000006
PS000006	445->449	CK2_PHOSPHO_SITE	PDOC000006
PS000008	12->18	MYRISTYL	PDOC000008
PS000008	65->71	MYRISTYL	PDOC000008
PS000008	76->82	MYRISTYL	PDOC000008
PS000008	193->199	MYRISTYL	PDOC000008
PS000008	267->273	MYRISTYL	PDOC000008
PS000008	311->317	MYRISTYL	PDOC000008
PS000008	336->342	MYRISTYL	PDOC000008
PS000008	339->345	MYRISTYL	PDOC000008
PS000008	353->359	MYRISTYL	PDOC000008
PS000008	368->374	MYRISTYL	PDOC000008
PS000008	373->379	MYRISTYL	PDOC000008
PS000008	435->441	MYRISTYL	PDOC000008
PS000008	461->467	MYRISTYL	PDOC000008
PS000008	490->496	MYRISTYL	PDOC000008
PS000008	494->500	MYRISTYL	PDOC000008
PS000013	122->133	PROKAR_LIPOPROTEIN	PDOC000013
PS000215	404->414	MITOCH_CARRIER	PDOC00189

## Pfam for DKF2phutet\_24j6.3

```

HMM_NAME      Laminin B (Domain IV)
HMM            *YWR1PERFLGDQvT=YGGkLe*
               Y+R  +  LG+++ + G + +
Query         538 YFRFAQNTLGKGLFACGPDAK      558

```

DKF2phutel\_2h3

-----

group: differentiation/development

DKF2phutel\_2h3 encodes a novel 267 amino acid protein, with similarity to ITM2 (integral membrane protein 2) of chicken and mouse.

The novel protein contains a prenyl group binding site (CAAX box) and seems to be post-translationally modified by the attachment of either a farnesyl or a geranyl-geranyl group. The similar gallus G. protein E25 a marker for chondro-osteogenic differentiation.

The new protein can find application as a useful marker for chondro-osteogenic cell differentiation and for the modulation of chondro-osteogenic cell differentiation.

strong similarity to mouse E25 and gallus E3-16

complete cDNA, EST hits  
complete cds according to E25 start at Bp 56  
putative transmembrane protein (1 TM)

Sequenced by AGOWA

Locus: unknown

Insert length: 2033 bp

Poly A stretch at pos. 2007, polyadenylation signal at pos. 1986

```
1 GGACCGAGGC TGCACCGGCA GAGGCTGCGG GCGGACGCG CGGGCCGGCG
51 CAGCCATGGT GAAGATTAGC TTCCAGCCCG CCGTGGCTGG CATCAAGGGC
101 GACAAGGCTG ACAAGGCGTC GCGCTCGGCC CCTGCGCCCG CCTCGGCCAC
151 CGAGATCCCTG CTGACGCCGG CTAGGGAGGA GCAGCCCCCA CAACATCGAT
201 CCAAGAGGGG GAGCTCAGTG GCGGCGTGT GCTACCTGTC GATGGGCATG
251 GTCGTGCTGC TCATGGGCCT CGTGTTCGCC TCTGTCTACA TCTACAGATA
301 CTTCTTTCTT GCACAGCTGG CCGAGATAA CTTCTCCCG TGTGGTGTGC
351 TGTATGAGGA CTCCCTGTCC TCCAGGTCC GGACTCAGAT GGAGCTGGAA
401 GAGGATGTGA AAATCTACCT CGACGAGAAC TACGAGCGCA TCAACGTGCC
451 TGTGCCCCAG TTGGCGGCG GTGACCCGTC AGACATCATC CATGACTTCC
501 AGCGGGGCTC GACTGCGTAC CATGATATCT CCTGGACAA GTGCTATGTC
551 ATCGAACTCA ACACCACCAT TGTGCTGCCC CCTCGCAACT TCTGGGAGCT
601 CCTCATGAAC GTGAAGAGGG GGACCTACCT GCCGCAGACG TACATCATCC
651 AGGAGGAGAT GGTGGTCACG GAGCATGTCA GTGACAAGGA GGCCCTGGGG
701 TCCTTCATCT ACCACCTGTG CAACGGGAAA GACACCTACC GGCTCCGGCG
751 CCGGGCAACG CGGAGGCGGA TCAACAAGCG TGGGCGCAAG AACTGCAATG
801 CCATCCGCCA CTTCGAGAAC ACCTTCGTGG TGGAGACGCT CATCTGCGGG
851 GTGGTGTGAG GCCCTCCTCC CCCAGAACC CCTGCGGTGT TCCTCTTTTC
901 TTTCTTCCAG CTGCTCTCTG GCCCTCCTCC TTCCCTCTGC TTAGCTTGTA
951 CTTTGGACGC GTTCTATAG AGGTGACATG TCTCTCCATT CCTCTCCAAC
1001 CTGCCCCACC TCCCTGTACC AGAGCTGTGA TCTCTCGGTG GGGGGCCCAT
1051 TCTGCTGAC CTGGGTGTGG CGGAGGGAGA GGCGATGCTG CAAAGTGTTC
1101 TCTGTGTCCC ACTGTCTTGA AGCTGGGCTT GCCAAAGCCT GGGCCACAG
1151 CTGACCGGCG AGCCCAAGGG GAAGGACCGG TTGGGGGAGC CGGGCATGTG
1201 AGGCCCTGGG CAAGGGGATG GGGCTGTGGG GGCGGGGCGG CATGGGCTTC
1251 AGAAGTATCT GCACAATTAG AAAAGTCCTC AGAAGCTTTT TCTTGGAGGG
1301 TACACTTTCT TCACTGTCCC TATTCCTAGA CCTGGGGCTT GAGCTGAGGA
1351 TGGACGATG TGCCAGGGA GGGACCCACC AGAGCACAAG AGAAGGTGGC
1401 TACCTGGGGG TGTCCAGGG ACTCTGTCAG TGCTTCAGC CCACGACAG
1451 GAGCTTGGAG TTTGGGAGT GGGGATGAGT CCGTCAAGCA CAACTGTTCT
1501 CTGAGTGGAA CCAAGAAGC AAGGAGCTAG GACCCCACT CCTGCCCCCT
1551 AGGAGCACAA GCAGGGTCCC CTCAGTCAAG GCAGTGGGAT GGGCGGCTGA
1601 GGAACGGGGC AGGCAAGGTC ACTGCTCAGT CACGTCCACG GGGGACGAGC
1651 CGTGGGTTCT GCTGAGTAGG TGGAGCTCAT TGCTTTCTCC AAGCTTGGAA
1701 CTGTTTGAAG AGATAACACA GAGGGAAGG GAGAGCCACC TGGTACTTGT
1751 CCACCCCTGCC TCCTCTGTTC TGAAATTCCA TCCCCTCAG CTTAGGGGAA
1801 TGCACCTTTT TCCCTTTCTT TCTCACTTTT GCATGTTTTT ACTGATCATT
1851 CGATATGCTA ACCGTTCTCA GCCCTGAGCC TTGGAGAGGA GGGCTGTAAC
1901 GCCTTCAGTC AGTCTCTGGG GATGAAACTC TTAATGCTT TGTATATTTT
1951 CTCATATTGA TCTCTTTTCA GAAGTGTCTA TAGAACATA AAAATCTTTT
2001 ACTTCTGAAA AAAAAAAAAA AAAAGGCGCG CCG
```

## BLAST Results

-----

Entry B64417 from database EMBL:  
CIT-HSP-2023A7.TR CIT-HSP Homo sapiens genomic clone 2023A7.  
Length = 715  
Plus Strand HSPs:

Score = 1546 (232.0 bits), Expect = 7.8e-64, P = 7.8e-64  
Identities = 310/311 (99%)

# Medline entries

96325063:  
Isolation of markers for chondro-osteogenic differentiation using cDNA library subtraction.  
Molecular cloning and characterization of a gene belonging to a novel multigene family of integral membrane proteins.

## Peptide information for frame 2

ORF from 56 bp to 856 bp; peptide length: 267  
Category: strong similarity to known protein

1 MVKISFQPAV AGIKGDKADK ASASAPAPAS ATEILLTPAR EEQPPQHRSK  
51 RGSSVGGVCY LSMGMVLLM GLVFASVYIY RYFFLAQLAR DNFFRCGVLY  
101 EDSLSSQVRT QMELEEDVKI YLDENYERIN VVPVQFGGDP ADIIHDFQR  
151 GLTAYHDISL DKCYVIELNT TIVLPPRNFW ELLMNVKRGT YLPQTYIIQE  
201 EMVVEHVSD KEALGSFIYH LCNGKDTYRL RRRATRRRIN KRGAKNCNAI  
251 RHFENTFVVE TLICGVV

## BLASTP hits

No BLASTP hits available

## Alert BLASTP hits for DKFZphut1\_2h3, frame 2

SWISSNEW:ITMB\_CHICK INTEGRAL MEMBRANE PROTEIN 2B (TRANSMEMBRANE PROTEIN E3-16)., N = 1, Score = 573, P = 1.3e-55

SWISSNEW:ITMB\_MOUSE INTEGRAL MEMBRANE PROTEIN 2B (E25B PROTEIN)., N = 1, Score = 560, P = 3.2e-54

SWISSNEW:ITMA\_HUMAN INTEGRAL MEMBRANE PROTEIN 2A (E25 PROTEIN)., N = 1, Score = 456, P = 3.3e-43

>SWISSNEW:ITMB\_CHICK INTEGRAL MEMBRANE PROTEIN 2B (TRANSMEMBRANE PROTEIN E3-16).

Length = 262

## HSPs:

Score = 573 (86.0 bits), Expect = 1.3e-55, P = 1.3e-55  
Identities = 117/264 (44%), Positives = 172/264 (65%)

Query: 1 MVKISFQPAVAGIKGDKADKASASAPASATEILLTPAREEQPPQHRSKRGSSVGGVCY 60  
MVK+SF A+A + A+K ++ ++L+ P ++P G  
Sbjct: 1 MVKVSFNSALA--HKEAANKKEENS-----QVLILPPDAKEPEDVVVPAGHKRAWCWC 51

Query: 61 LSMGMVLLMGLVFASVYIYRYFFLAQLARDNFFRCGVLY-EDSL-----SQVRTQM-- 112  
+ G+ +L G++ Y+Y+YF Q + CG+ Y ED LS +Q+++  
Sbjct: 52 MCFGLAFMLAGVILGGAYLYKYFAFQ---GGVYFCGIKIEDGLSLPESGAQLKSARYH 108

Query: 113 ELEDVKIYLDENYERINVPVQFGGDPADIHDFQRLTAYHDISLDKCYVIELNTTI 172  
+E+++I +E+ E I+VVPV+F DPADI+HDF R LTAY D+SLDKCYVI LNT++  
Sbjct: 109 TIEQNIQILEEEDVEFISVPVPEFADSDPADIHDFHRLTAYLDLSLDKCYVIPLNTSV 168

Query: 173 VLPFRNFWELLMNVKRGTLYLPQTYIIQEEMVVEHVSDKEALGSFIYHLCNGKDTYRLRR 232  
V+PP+NF ELL+N+K GTYLPQ+Y+I E+M+VT+ + + + LG FIY LC GK+TY+L+R  
Sbjct: 169 VMPPKNFLELLINIKAGTYLPQSYLIHEQMIVTDRIENVQLGFFFIYRLCRGKETYKLQR 228

Query: 233 RATRRRINKRGAKNCNAIRHFENTFVVETLIC 264  
+ + I KR A NC IRHFEN F +ETLIC  
Sbjct: 229 KEAMKGIQKREAVNCRKIRHFENRFAMETLIC 260

## Pedant information for DKFZphut1\_2h3, frame 2

## Report for DKFZphut1\_2h3.2

[LENGTH] 267  
 [MW] 30253.96  
 [pI] 8.16  
 [HOMOL] SWISSNEW:ITMB\_CHICK INTEGRAL MEMBRANE PROTEIN 2B (TRANSMEMBRANE PROTEIN E3-16).  
 1e-49  
 [PROSITE] MYRISTYL 4  
 [PROSITE] PRENYLATION 1  
 [PROSITE] CAMP\_PHOSPHO\_SITE 3  
 [PROSITE] CK2\_PHOSPHO\_SITE 3  
 [PROSITE] TYR\_PHOSPHO\_SITE 1  
 [PROSITE] PKC\_PHOSPHO\_SITE 4  
 [PROSITE] ASN\_GLYCOSYLATION 1  
 [KW] TRANSMEMBRANE 1  
 [KW] LOW\_COMPLEXITY 15.36 %

SEQ MVKISFQPAVAGIKGDKADKASASAPASATEILLTPAREEQPPQHRSKRGSSVGGVCY  
 SEG .....XXXXXXXXXXXXXXXXX.....  
 PRD cccccccchhhhhhhhhhhhhhhcccccceccccccccccccccccccchh  
 MEM .....MMM  
  
 SEQ LSMGMVVLMLGVFASVYIYRYFFLAQLARDNFFRCGLVYEDSLSSQVRTQMELEEDVKI  
 SEG ..XXXXXXXXXXXXX.....  
 PRD hhhhhhhhhhhhhhhhhhhhhcchhhhhhhhhhhccceccccccccchhhhhhhhhhh  
 MEM MMMMMMMMMMMMMMMMMMMMM  
  
 SEQ YLDENYERINVPVPQFGGDPADIHDFQRGLTAYHDISLDKCYVIELNTTIVLPPRNFV  
 SEG .....  
 PRD hhccccccccccccccccchhhhhhhhhhhhhccceccccccccchh  
 MEM .....  
  
 SEQ ELLMNVKRGTYLPQTYIIQEEMVTEHVSDKEALGSFIYHLCNGKDTYRLRRRATRRRIN  
 SEG .....XXXXXXXXXXXXX.....  
 PRD hhhhhccccccccceehhhhhhhccccchhhhhheccccchhhhhhhhhhhhh  
 MEM .....  
  
 SEQ KRGAKNCNAIRHFENTFVVETLICGVV  
 SEG xx.....  
 PRD hhhccccccccchhhhhhecccc  
 MEM .....

## Prosites for DKFZphut1\_2h3.2

PS00001	169->173	ASN_GLYCOSYLATION	PDOC00001
PS00004	50->54	CAMP_PHOSPHO_SITE	PDOC00004
PS00004	187->191	CAMP_PHOSPHO_SITE	PDOC00004
PS00004	232->236	CAMP_PHOSPHO_SITE	PDOC00004
PS00005	49->52	PKC_PHOSPHO_SITE	PDOC00005
PS00005	209->212	PKC_PHOSPHO_SITE	PDOC00005
PS00005	227->230	PKC_PHOSPHO_SITE	PDOC00005
PS00005	235->238	PKC_PHOSPHO_SITE	PDOC00005
PS00006	30->34	CK2_PHOSPHO_SITE	PDOC00006
PS00006	110->114	CK2_PHOSPHO_SITE	PDOC00006
PS00006	209->213	CK2_PHOSPHO_SITE	PDOC00006
PS00007	119->127	TYR_PHOSPHO_SITE	PDOC00007
PS00008	52->58	MYRISTYL	PDOC00008
PS00008	71->77	MYRISTYL	PDOC00008
PS00008	138->144	MYRISTYL	PDOC00008
PS00008	243->249	MYRISTYL	PDOC00008
PS00294	264->268	PENYLATION	PDOC00266

(No Pfam data available for DKFZphut1\_2h3.2)

DKFZphmcfl\_1a11

-----

group: transmembrane protein

DKFZphmcfl\_1a11 encodes a novel 393 amino acid protein with weak similarity to S.pombe SPBC29A3\_3 protein and S. cerevisiae putative membrane protein YDR255c.

The novel protein contains 1 transmembrane region.  
No informative BLAST results; No predictive prosite, pfam or SCOP motive.

The new protein can find application in studying the expression profile of mammary carcinoma-specific genes and as a new marker for mammary carcinoma cells.

similarity to YDR255c and SPBC29A3.03c

membrane regions: 1

Summary DKFZphmcfl\_1a11 encodes a novel 393 amino acid protein, with  
similarity to YDR255c and SPBC29A3.03c.

similarity to YDR255c and SPBC29A3.03c

complete cDNA, complete cds, EST hits

potential start at Bp 110 matches kozak consensus

Sequenced by DKFZ

Locus: /map="542.7 cR from top of Chr5 linkage group"

Insert length: 1819 bp

Poly A stretch at pos. 1808, no polyadenylation signal found

```

1  CCCGCCCCAG  CCCCCGAAGA  GCCGCCTCAG  CCGGGGGGAG  TTGCTCGGAC
51  TCAAACTGTC  AGTCCTCGTG  CGACCGCGCT  GGGTCGGAAG  TGAGCAGGCT
101  GAGGCCACCA  TGGAGCAGTG  TGCCTGCCGTG  GAGAGAGAGC  TGGACAAGGT
151  CCTGCAGAAG  TTCCTGACCT  ACGGGCAGCA  CTGTGAGCGG  AGCCTGGAGG
201  AGCTGTGCA  CTACGTGGGC  CAGCTGCCGG  CTGAGCTGGC  CAGCGCAGCC
251  CTCACGGGGA  CCCCTCTCTC  AGCCACCCCT  TCTCTGGTGA  TGTACAGTG
301  CTGCCGGAAG  ATCAAGATA  CGGTGCAGAA  ACTGGCTTCG  GACCATAAGG
351  ACATTCACAG  CAGTGTATCC  CGAGTGGGCA  AAGCCATTGA  CAGGAACCTC
401  GACTCTGAGA  TCTGTGGTGT  TGTGTGAGAT  GCGGTGTGGG  ACGCGCGGGA
451  ACAGCAGCAG  CAGATCCTGC  AGATGGCCAT  CGTGGAAACAC  CTGTATCAGC
501  AGGCGATGCT  CAGCGTGGCC  GAGGAGCTGT  GCCAGGAATC  AACGCTGAAT
551  GTGGACTTGG  ATTTCAAGCA  GCCTTTCCTA  GAGTTGAATC  GAATCTGGA
601  AGCCCTGCAC  GAACAAGACC  TGGGTCTGCG  GTTGAATGG  GCCGTCTCCC
651  ACAGGCGAGC  CCGTCTGGAA  CTCAACAGCT  CCCTGGAGTT  CAAGCTGCAC
701  CGACTGCAC  TCATCCGCT  CTTGGCAGGA  GGCCCCGCGA  AGCAGCTGGA
751  GGCCCTCAGC  TATGCTCGGC  ACTTCCAGCC  CTTTGCTCGG  CTGCACCAGC
801  GGGGATCCCA  GGTGATGATG  GGCAGCCTGG  TGTACCTCGG  GCTGGGCTTG
851  GAGAAGTCAC  CCTACTGCCA  CCTGTGGAG  AGCAGCCACT  GGGCAGAGAT
901  CTGTGAGACC  TTTACCCGGG  ACGCCTGTT  CCTGTGGGG  CTTTCTGTGG
951  AGTCCCCCT  TAGCGTCAGC  TTTGCCTCTG  GCTGTGTGGC  GCTGCCTGTG
1001  TTGATGAACA  TCAAGGCTGT  GATTGAGCAG  CGGCAGTGCA  CTGGGGTCTG
1051  GAATCAACA  GACGAGTTAC  CGATTGAGAT  TGAACTAGGC  ATGAAGTGCT
1101  GGTACCAC  CGTGTTCGCT  TGCCCATCC  TCCGCCAGCA  GACGTCAGAT
1151  TCCAACCTC  CCATCAAGCT  CATCTGTGGC  CATGTTATCT  CCCGAGATGC
1201  ACTCAATAAG  CTCATTAATG  GAGGAAAGCT  GAAGTGTCCC  TACTGTCCCA
1251  TGGAGCAGAA  CCCGGCAGAT  GGGAAACGCA  TCATATTCTG  ATTCTACCT
1301  GGAAGGAATT  TTGTTGAAAG  GGGTTTTCAC  CTGTGAGCCT  TGGTCTGTCT
1351  CGGTAGGGTG  GTCAACTTCA  GTGGACTGTG  GTTGGTTTCA  GAGCGCCTGG
1401  CTGAGGAGTT  CCACTGAGGG  GAGCACTGGA  GCAGCCCTTT  GGCAGAGGCT
1451  GAGGAGGGAG  ATGGACCAGC  CCACGCTGG  CACCTGGCTC  CATGGCATAA
1501  GGAAGGGGAG  ATGCTGGCCT  CTGTGCTCCT  GCTGTCTTTT  CCTGTTTCTG
1551  TTTGCGTTTG  ACTTAGTAGC  AACCGACAGA  GTGGCAAGGG  ATTTGGTCTT
1601  CAGCAGTAGA  CATCTTCCA  CCCCTGCCCT  CAGCCAAGTC  TCTGTCTGCC
1651  ATGCCAATGC  TATGTCACC  CTTGCCCTC  GGCCCAAGAG  TGTCCAGCGG
1701  TGGCCACCT  TTCTCTCCA  CTACAGCCTC  AACAGTATGT  ACCATCTCCC
1751  ACTGTAAATA  GTCCAGTTA  GAACGGAAATG  CCGTGTGTTT  ATAACTTTGA
1801  ACAATGTAA  AAAAAAAAA

```

## BLAST Results

-----

Entry HS579359 from database EMBL:  
human STS WI-6350.  
Score = 1027, P = 9.9e-40, identities = 207/209

## Medline entries

No Medline entry

## Peptide information for frame 2

ORF from 110 bp to 1288 bp; peptide length: 393  
 Category: similarity to unknown protein

```

1 MEQCACVERE LDKVLQKFLT YGQHCERSLE ELLHYVGQLR AELASAAALQG
51 TPLSATLSLV MSQCCRKIKD TVQKLASDHK DIHSSVSRVG KAIDRNFDSE
101 ICGVSDAVW DAREQQQIL QMAIVEHLYQ QGMLSVAEEL CQESTLNVDL
151 DFKQPFLELN RILEALHEQD LGPALEWAVS HRQRLLELNS SLEFKLHRLH
201 FIRLLAGGPA KQLEALSYAR HFQPFARLHQ REIQVMGSL VYLRGLGLEKS
251 PYCHLLDSSH WAEICETFTF DACSLGLSV ESPLSVSFAS GCVALPVLMMN
301 IKAVIEQRQC TGVWNHKKDEL PIEIELGMKC WYHSVFACPI LRQQTSDSNP
351 PIKLCIGHVI SRDALNKLIN GGLKCPYCP MEQNPADGKR IIF

```

## BLASTP hits

No BLASTP hits available

## Alert BLASTP hits for DKFZphmcf1\_lal1, frame 2

TREMBL:SPBC29A3\_3 gene: "SPBC29A3.03c"; product: "hypothetical protein"; S.pombe chromosome II cosmid c29A3., N = 2, Score = 302, P = 3.4e-42

PIR:S67312 probable membrane protein YDR255c - yeast (Saccharomyces cerevisiae), N = 1, Score = 271, P = 5.3e-22

TREMBL:CET07D1\_2 gene: "T07D1.2"; Caenorhabditis elegans cosmid T07D1., N = 1, Score = 193, P = 5.6e-13

>TREMBL:SPBC29A3\_3 gene: "SPBC29A3.03c"; product: "hypothetical protein"; S.pombe chromosome II cosmid c29A3. Length = 398

## HSPs:

Score = 302 (45.3 bits), Expect = 3.4e-42, Sum P(2) = 3.4e-42  
 Identities = 55/142 (38%), Positives = 89/142 (62%)

Query: 252 YCHLLDSSHWAEICETFTRDACSLGLSVESPLSVSFASGCVALPVLMMNIKAVIEQRQCT 311  
 Y +LD W + F R+ C+ LG+S+ESPL + +G +ALP+L+ + +++++  
 Sbjct: 258 YIDVLDLD-WKSLELLFVREFCAALGMSLESPLDIVVNAGAIAPILLKMSSIMKKKHT 316

Query: 312 GVNHHKDELPIEIELGMKCWYHSVFACPIRQQTSDSNPPIKLCIGHVISRDALNKLING 371  
 W + ELP+EI L +HSVF CP+ ++Q ++ NPP+ + CGHVI ++L +L  
 Sbjct: 317 --WTSQGELPVEIFLPSSYHFHSVFTCPVSKEQATEENPPMMMSCGHVIVKESLRQLSRN 374

Query: 372 G--KLKCPYCPMEQNPADGKRIF 393  
 G + KCPYCP E AD R+ F  
 Sbjct: 375 GSQRFKCPYCPNENVAADAIRVF 398

Score = 161 (24.2 bits), Expect = 3.4e-42, Sum P(2) = 3.4e-42  
 Identities = 51/221 (23%), Positives = 102/221 (46%)

Query: 22 GQHCERSLEELLHYVGQLRAELASAAALQGTPLSATLSLVMSQCCRKIKD TVQKLASDHKD 81  
 G C L EL + + + L+ P++ LV C K + L K  
 Sbjct: 15 GNKCLAKLNL---ESILKDAKKSCLKD-PTSMKELVA--CSEKTTQVFDLKRTEKK 67

Query: 82 IHSSVSRVGKAIDRNFDSEICGVVSDAVWDAREQQQILQMAIVEHLYQQGMLSVAEELC 141  
 H+S++R GK +++ F+ ++ + ++++++ + A+ H ++QG + +A C  
 Sbjct: 68 FHTSLNRFGKTLEKKFNFDLEDIKLHSSFESKKRE---IDTALSLHFFRQGDVLAHLFC 124

Query: 142 QESTLNVDLDFKQPFLELNRIEALHEQDLGPALEWAVSHRQRLLELNSLEFKLHRLHF 201  
 +E+ + + F L I++ ++DL +EWA R L SSLE+ L +  
 Sbjct: 125 KEAGIEEPSSELHVFTLLKSIVQGIRDKOLKLPIEWASQCRGYLERKGSSELYTLQKYRL 184

Query: 202 IRLLAGGPAQL-EALSYAR-HFQPFARLHQREIQVMGSLVY 242  
 + K + A+ Y R + F + H +IQ M +L +



Report for DKFZphmcf1 1a11.2

[illegible]

PS00001	189->193	ASN_GLYCOSYLATION	PDOC00001
PS00005	180->183	PKC_PHOSPHO_SITE	PDOC00005
PS00006	28->32	CK2_PHOSPHO_SITE	PDOC00006
PS00006	135->139	CK2_PHOSPHO_SITE	PDOC00006
PS00006	190->194	CK2_PHOSPHO_SITE	PDOC00006
PS00007	211->219	TYR_PHOSPHO_SITE	PDOC00007
PS00007	27->36	TYR_PHOSPHO_SITE	PDOC00007
PS00007	244->253	TYR_PHOSPHO_SITE	PDOC00007
PS00008	37->43	MYRISTYL	PDOC00008
PS00008	50->56	MYRISTYL	PDOC00008
PS00009	387->391	AMIDATION	PDOC00009
PS00013	282->293	PROKAR_LIPOPROTEIN	PDOC00013

545

DKF2phmcf1\_1c23

-----

group: mammary carcinoma derived

DKF2phmcf1\_1c23.1 encodes a novel 311 amino acid proline rich protein.

No informative BLAST results; No predictive prosite, pfam or SCOP motif.

The new protein can find application in studying the expression profile of mamma carcinoma-specific genes.

unknown, proline rich protein

complete cDNA, complete cds? potential start at Bp 50, EST hits

Sequenced by DKFZ

Locus: unknown

Insert length: 3077 bp

Poly A stretch at pos. 3067, polyadenylation signal at pos. 3048

```
1  AACTGGCCCC CTCCCCACC CCCTGCCCCT GAGGAGCAGG ACCTGTCCAT
51  GGCTGACTTC CCCCCACCAG AGGAGGCTTT TTCTCTGTG GCCAGCCCTG
101 AGCCTGCAGG CCCTTCAGGC TCCCAGAGC TTGTACAGTC CCGGCTGCT
151 TCGTCTCTCT CAGCTACTGC TTTCAGATT CAGCCCCCGG GTAGCCAGA
201 CCCTCTCTCA GCTCCGCCAG CCCCAGCTCC TGCTAGTTCC GCCCCAGGGC
251 ATGTGGCCAA GCTCCCTCAG AAGGAACCGG TGGGCTGTAG CAAGGGTGGT
301 GGGCTTCCCA GGGAGGACGT AGGTGCGCCC CTGGTCACGC CCTCGCTCCT
351 GCAGATGGTG CGGCTGCGCT CCGTGGGTGC TCCAGGAGGG GCTCCACCCC
401 CAGCACTGGG GCCATCGGCC CCCCAGAAAC CACTGCGAAG GGCCTGTCTA
451 GGGCGGGCCA GCCCAGTGCC TGCCCCCTCC TCAGGGCTCC ATGCTGCGGT
501 CCGACTCAAG GCCTGCAGCC TGGCCGCCAG TGAAGGCTC TCAAGTGCTC
551 AGCCCAACGG ACCGCTGAG GCAGAGCCAC GGCTCCCCA GTCCCCTGCC
601 TCAACGGCCA GTTTCATCTT CTCCAAGGGC TCTAGGAAGC TGCAGCTGGA
651 GCGGCCCGTG TCCCCTGAGA CCCAGGCTGA CCTCCAGCGG AATCTGGTGG
701 CAGAACTCCG GAGCATCTCA GAGCAGCGGC CACCCAGGC CCAAGAAG
751 TCACCTAAGG CTCCCCACC TGTGGCCCGC AAGCCGTCTG TGGGAGTCCC
801 CCCACCCGCG TCCCCAGTT ACCCTCGAGC TGAGCCCTTT ACTGCTCCTC
851 CCACCAATGG GCTCCCTCAC ACCCAGGACA GGAATAAGAG GGAGCTGGCG
901 GAGAAATGGAG GTGTCTGCA GCTGGTGGGC CCAGAGGAGA AGATGGGCT
951 CCGGGGCTCA GACTCAGAGA AAGAGCTGGC CTGACCACCA GGCACCTCAC
1001 TGGCACTGCT GACCCATCCC AGAAACACAA TCTCAGGGAC CCGAGCAGCT
1051 CCAAGGACGA GAGGATACAG CAGACACAAC CTAATAGAGA GGGCGCTGCG
1101 AGCCTTAACG TCCACGGCCT TCGATACCTA TGCAAGCCTG GTGTGTCTCC
1151 TGTCTCTAGA GTCATCTGCG GCTCATGCCT TTCCCGAAT GGGTTACCTT
1201 CTGGCAGTTG CCGCTTCAGT CTTGGCCTTA GCCTCATCTT GAAGTGGGTA
1251 GCTGGCGGGA GAGGGTGGCT GCGCCCTGCT CTGGCCCTGA GGCTGCAGAG
1301 TTGGGAGCAG GACACCTCAC CTGAGTTTCA TTTTTTTTCA TGTCCAAACC
1351 ATGCACATAC TATAGTCCAG AATCAAAGCA CTTTTGAAAA GTGGCTGCAT
1401 GGGCATCCTC CAGGGCCAGG GAAGTTGCAT TCCAAGGGCC TGTTTACATG
1451 GCAGCAGAAT CCATCCCCGG CAGTCAGCCC ATAGCTTGGG ACCAGTCTGT
1501 GCCCTCTCTG CCAGTCCAGT TTAATCTCTT TGCTTCCTGA AGGTGGCCAA
1551 GTCATTGTGT TCCACAGGCG TTCTTAGGCG TGGGGGAGG TGTGGGGCTG
1601 TGGAAATCCA AAGCACAAAA GGTGCAGAGG GGATTGGCCT TCCTGTGCCT
1651 CAACTCACCA ACCACCTCCG TGCTTCCAG TTCTGCCAGG TGCTCCATGC
1701 TGGGGACAAG TAGGAGACTG CCAGGGCCCA AAGAAATGGG TGAGCAGTAG
1751 AGTCATCTCG GGGCACTTGG CAGTGTCAAG CACCTGCCCC TTGCCTCCTT
1801 GACCACACTG GGGTGGGTGG GCCCCAGCA CTTCAGAGGC AGGAGCCTTT
1851 GGGCTGAGCA AGCACTGAGG AGGTGGATGG AAGGGAGCAT CTGGAGGGGG
1901 GGAGCTTCTT TGAGCAGTGG GCCCAGGCCT GGCCCTCCAC ACTTCATTCT
1951 CTGACCTTTC TCTCTCTTCA TTTCCGTGCA TGTCTTTCT GCAGCTGCCT
2001 TTCAGCACAG GTGGTTCCAC TGGGGGCGAG TAACGCTGAG TGACAAGGAT
2051 GGGAAAGCCAC AGGTGCATTT TACTCAAGTC TTCTCTAGTC AATGAGGGGC
2101 ACCCAGTGCT TTAGGGCAG GCTGGGTGGT GGTCCCCTAG GTATCAGCCT
2151 CTCTTACTGT ACTCTCCGGG AATGTTAACC TTTCTATTTT CAGCCTGTGC
2201 CAGCTGTCTA GGAAGCTGG CTTCGCCATT GGCCCTGTGT GGTCCACAGC
2251 AGCGTGGCTG CCCCCAGGG CCACCGCTTC TTTCTTGATC CTCTTTCCTT
2301 AACAGTGACT TGGGCTTGAG TCTGGCAAGG AACCTTGCTT TTAGCTTCAC
2351 CACCAAGGAG AGAGGTTGAC ATGACCTCCC CGCCCCCTCA CCAAGGCTGG
2401 GAACAGAGGG GATGTGGTGA GAGCAGGTT CCTCTGGCCC TCTCCAGGGT
2451 GTTTTCCACT AGTCACTACT GTCTTCTCCT TGTAGCTAAT CAATCAATAT
2501 TCTTCCCTTG CCTGTGGGCA GTGGAGAGTG CTGCTGGGTG TACGCTGCAC
2551 CTGCCCAGTG AGTTGGGGAA AGAGGATAAT CAGTGAGCAC TGTCTGCCTC
2601 AGAGCTCTCT ATCTACCCCA CCCCCTAGGA TCCAGGACTG GGTCAAAGCT
2651 GCATGAAACC AGGCCCTGGC AGCAACCTGG GAATGGCTGG AGGTGGGAGA
2701 GAACCTGACT TCTCTTTCCC TCTCCCTCCT CCAACATTAC TGGAACTCTA
```

```

2751 TCCTGTTAGG ATCTTCTGAG CTTGTTTCCC TGCTGGGTGG GACAGAGGAC
2801 AAAGGAGAAG GGAGGGTCTA GAAGAGGCAG CCCTTCTTTG TCCTCTGGGG
2851 TAAATGAGCT TGACCTAGAG TAAATGGAGA GACCAAAAGC CTCTGATTTT
2901 TAATTTCCAT AAAATGTTAG AACTATATAT ATACATATAT ATATTTCTTT
2951 AAATTTTGA GTCTTTGATA TGTCTAAAAA TCCATTCCCT CTGCCCTGAA
3001 GCCTGAGTGA GACACATGAA GAAACTGTG TTTTATTAA AGATGTTAAT
3051 TAAATGATTG AAACCTGAAA AAAAAAA

```

## BLAST Results

No BLAST result

## Medline entries

No Medline entry

## Peptide information for frame 1

ORF from 49 bp to 981 bp; peptide length: 311  
 Category: putative protein  
 Classification: unset

```

1 MADFPPEEA FFSVASPEPA GPSGSPPELV SPAASSSSAT ALQIQPPGSP
51 DPPPPAPPAPA PASSAPGHVA KLPOKEPVGC SKGGGPPRED VGAPLVTPSL
101 LQMVRLRSVG APGGAPTAL GPSAPQKPLR RALSGRASP PAPSSGLHAA
151 VRLKACSLAA SEGLSSAQPN GPPEAEPRPP QSPASTASFI FSKGSRKLQL
201 ERVSPETQA DLQRNLVAEL RSISEQRPPQ APKKSPPKAP PVARKPSVGV
251 PPPASPSYPR AEPLTAPPTN GLPHTQDRTK RELAENGVL QLVGPEEKMG
301 LPGSDSQKEL A

```

## BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphmcf1\_1c23, frame 1

PIR:S49915 extensin-like protein - maize, N = 1, Score = 215, P = 6.1e-15

PIR:A28996 proline-rich protein M14 precursor - mouse, N = 1, Score = 191, P = 3.8e-13

>PIR:S49915 extensin-like protein - maize  
 Length = 1,188

## HSPs:

Score = 215 (32.3 bits), Expect = 6.1e-15, P = 6.1e-15  
 Identities = 81/269 (30%), Positives = 115/269 (42%)

```

Query:      5 PPPEEAFFS----VASPEPAGPSGSPPELVSSPAASSSSATALQIQPPGSP--DPPP---A 55
           PPP  S  V SP P P SP  PA +SS ++ PP +P PPP  +
Sbjct:    598 PPPPAPVASPPPPVKSPPTPVASPP---PPAPVASSPPPMKSPPTPVSSPPPEKS 654

Query:      56 PPAPAPASSAPGHVAKLPQKEPVGC SKGGGPPREDVGCAPLVTPSLLQMVRLRSVGAPGGA 115
           PP P PA S P  + P  P  K PP ++ P + PS  +  P
Sbjct:    655 PPPPPPAKSTPPP-EEYPT--PPTSVKSSPPPEKSLPPPTLIPSPPPQEKPTPPSTPSKP 711

Query:     116 PTPALGPSAPQKPLRRA-LSGRASVPVPAPSSGLHAAVRLKACSLAASEGLSSAQFNGPPE 174
           P+  PS P++P+  + ++SP PAP S  +LA  S + + PP
Sbjct:    712 PSSPEKPSPPKEPVSSPPQTPKSSPPAPVSSPPPTPVSSPPALAPVSSPPSVKSSPPPA 771

Query:     175 AEP RPPQSPASTASFI FSKGSRKLQLERPV-SPETQADLQRNLVAELRSISEQRPPQAPK 233
           PP +P  +S  +Q+ P +P++ L  V+  + + PP AP
Sbjct:    772 PLSSPPAPQVKSS-----PPPQVSSPPAPKSSPPLAP--VSSPPQVKTSPPPPAPL 823

Query:     234 KSPKAPPPVARKPSVGV--PPPASPSYPRAEPLTAPPTNGLP 273
           SP  P  + P V V PPP  S P  P+++PP  P
Sbjct:    824 SSPPLAPK-SSPPHVVSPPPPVVKSSPPAPVSSPPLTPK 864

Score = 206 (30.9 bits), Expect = 9.1e-14, P = 9.1e-14

```

Identities = 82/261 (31%), Positives = 108/261 (41%)

Query: 17 PEPAG-PSGSPVLSSPAASS---SSATALQIQPGSPDPPFPAP---PAPAPASSAPGHV 69  
 P P G P SP + PAAS+ S T + P P+P P P P P P +P  
 Sbjct: 410 PTPGGGPPSSP-VPGKPAASAPMPSPHTPPDVSPLEPLPEPSVPAPAPMPMPTPHSPPAD 468

Query: 70 AKLPQKEPV-GCSKGGGPPREDVGAPLVTPLSLQMVRLRSVGAPGGAPTALGSPAPQKP 128  
 +P PV G S P V P + +V+L AP G+P P + ++P P  
 Sbjct: 469 DYPPTTPVPGKSPSPATSPSPQVQPPAASTPPPSLVKLSPPQAPVGSPPPPVKTTSPPAP 528

Query: 129 LRRALSGRASVPAPSSGLHAAVRLKACSLAASEGLSSAQNGPPEAEPRPPQSPASTAS 188  
 + G SP P P S + +K+ A G + P PPE P PP AS  
 Sbjct: 529 I-----GSPSP-PPPVSVSPPPPVKSPPPAPVG---SPP--PPEKSPPPAPVASPPP 577

Query: 189 FIFSKGRKLQLERPVS PETQADLQRLNVAELRSISEQRPPQAPKKS PKAPPPVARKPS- 247  
 + S L P P ++ VA + PP P SP P PVA P  
 Sbjct: 578 PVKSPPPPTLVASPP--PPVKSPPPAPVASPPPVKSPPPPTPVASPPPPAPVASSPPP 635

Query: 248 VGVPPP---ASPSYPRAEPLTAPNTNGLPHTQD 277  
 + PPP +SP P P PP P ++  
 Sbjct: 636 MKSPPTPVSSPPPEKSPPPPPAKSTPPPEE 669

Score = 202 (30.3 bits), Expect = 2.9e-13, P = 2.9e-13  
 Identities = 81/254 (31%), Positives = 110/254 (43%)

Query: 16 SPEPAGPSGSPVLSSP---AASSSSATALQIQPGSP-DPPAPPAPAPASSAPGHVA 70  
 SP PA P SP L SSP SS ++ PP +P PP P PA S P HV+  
 Sbjct: 817 SPPFA-PLSSPPLAPKSSPPHVVSPPPVKSSPPAPVSSPPLTPKPA---SPPAHVS 872

Query: 71 KLPQ---KEPVGCSKGGGPPREDVGAPLVTPLSLQMVRLRSVGAPGGAPTALGSPAPQ 126  
 P+ P + PP E +P TP L ++S P +P +P +  
 Sbjct: 873 SPPEVVKPSTPPAPTIVISPESEPKSPPTPVSLPPPIVKSSPPAMVSSPMTKSP 932

Query: 127 KPLRRAL---SGRASVPAPSSGLHAAVRLKACSLAASEGLSSAQNGPPEAEPRPPQSP 183  
 P+ + + ++SP PAP S A K+ A L P PPE + PP +P  
 Sbjct: 933 PPVVSSPPPTVKSSPPAPVSSPATP--KSSPPAPVNL---P--PPEVKSSPPPTP 984

Query: 184 ASTASFIFSKGRKLQLERPVS PETQADLQRLNVAELRSISEQRPPQAPKKS PKAPPPVA 243  
 S+ + P PE ++ V+ + PP AP SP PPV  
 Sbjct: 985 VSSPPAPKSSPPAPMSSPPPEVKSPPPAPVSSPPPVKSPPPAPVSSP---PPPVK 1042

Query: 244 RKPS---VGVPPPASPSPYRAEPLTAPP 268  
 P V PPP S P P++PP  
 Sbjct: 1043 SPPPPAPVSSPPPVKSPPPAPISSP 1070

Score = 190 (28.5 bits), Expect = 7.9e-12, P = 7.9e-12  
 Identities = 74/264 (28%), Positives = 111/264 (42%)

Query: 5 PPPEEAFFSVASPEPAGPSGSPVLSSPAAS-SSSATALQIQPGSPDPPFPAPAPAS 63  
 PPP S PE + P P + P + T+++ PP PP P+P  
 Sbjct: 639 PPPPTPVSSPPPEKSPPPPPAKSTPPPEEYPTPTSVKSSPPPEKSLPPPTLIPSP 698

Query: 64 SAPGHVAKLPQKEPVGCSKGGGPPREDVGAPLVTPLSLQMVRLRSVGAPGGAPTALGSP 123  
 P K P K PP+E V +P TP V +P PTP P  
 Sbjct: 699 QEKTPTPTPSKPPSSPEKPS-PPKEPVSSPPQTPK--SSPPAPVSSP--PPTPVSSP 753

Query: 124 APQKPLRRALSGRASVPAPSSGLHAAVRLKACSLAASEGLSSAQNGPPEAEPRPPQSP 183  
 A P+ S ++SP PAP S A ++K+ + + P PP + PP +P  
 Sbjct: 754 A-LAPVSSPPSVKSSPPAPLSSPPAPQVKS---SPPVQVSSP--PPAPKSSPLAP 806

Query: 184 ASTASFIFSKGRKLQLERP-VSPETQADLQRLNVAELRSISEQRPPQAPKKS PKAPPPV 242  
 S+ + L P ++P++ +V+ + + PP AP SP P  
 Sbjct: 807 VSSPPQVEKTSPPAPLSSPPLAPKSSPP--HVVSSPPPVKSSPPAPVSSPPLTPKP 864

Query: 243 ARKPS-VGVPPP---PASPSYPR-----AEPLTAPP 268  
 A P+ V PP P++P P +EP ++PP  
 Sbjct: 865 ASPPAHVSSPPEVVKPSTPPAPTIVISPPSEPKSSP 901

Score = 189 (28.4 bits), Expect = 1.0e-11, P = 1.0e-11  
 Identities = 86/271 (31%), Positives = 112/271 (41%)

Query: 5 PPPEEAFFSVASPEPAGPSGSPVLSSP---AASSSSATALQIQPG--SPDPPFPAP--- 56  
 PPP A S P P S P + VSSP A SS A PP PPPAP  
 Sbjct: 768 PPP--APLSSPPAPQVKSPPPVQVSSPPAPKSSPLAPVSSPPQVEKTSPPAPLSS 825

Query: 57 PAPAPASSAPGHVAKLPQKEPVGCSKGGGPPREDVGAPLVTPLSLQMVRLRSVGAPGGAP 116  
 P AP SS P V P PV S PP V +P +TP V +P  
 Sbjct: 826 PPLAPKSSPPHVVSPP--PVVKSS---PPAPVSSPPLTPKPAFPA--HVSSPPEVV 878

Query: 117 TPALGSPAPQKPLRRALSGRASVPAPSSGLHAAVRLKAC-SLAASEGL---SSAQ-- 169  
 P+ P AP + ++SP P P S V+ ++ +S + SS P

Sbjct: 879 KPST-PPAPTTVISPPSEPKSSPPPTPVSLPPPIVKSSPPPPAMVSSPPMTPKSSPPPVV 937

Query: 170 -NGPPEAEPRPPQSPASTASFIFSKGSRKLQLERPVSPETQADLQRLVAELRSISEQRP 228  
+ PP + PP + P S + + P PE ++ V+ + P

Sbjct: 938 SSPPPTVKSSPPAPVSSPPATPKSSPPAPVNL-PPEVKSSPPPTPVSSPPPAKSSP 996

Query: 229 PQAPKKSAPKPPVARKPS---VGVPASPSPYPRAEPLTAPP 268  
P AP SP PPP + P V PPP S P P+++PP

Sbjct: 997 PPAPMSSP--PPEVKSSPPAPVSSPPPVKSSPPAPVSSPP 1038

Score = 181 (27.2 bits), Expect = 8.8e-11, P = 8.8e-11  
Identities = 73/277 (26%), Positives = 105/277 (37%)

Query: 3 DFPPEEAEFFSVASPEPAGPSGSELVSSPAASSSATALQIQPP---GSPDPP---PA 55  
D+ PP V P S SP+ V PAAS+ + +++ PP GSP PP +

Sbjct: 469 DYVPPTTP---VPGKSPATSPSQ-VQPPAASTPPSLVKLSPPQAPVGSPPPVKTT 524

Query: 56 PPAPASSAPGHVAKL---PQKEPVGCSKGGPPREDVGAPLVTPSLQMVRLRSVGA 111  
PPAP + S P V+ + P K P + G PP + P P ++S

Sbjct: 525 PPAPIGSPSPPPVSVSSPPPVKSSPPAPVGSPPPEKSPPPAPVASPPPVKSSPP 584

Query: 112 PG--GAPTALGPSAPQKPLRRA---LSGRASPVAPSSGLHAAVRLKACSLAASEGLSS 166  
P +P P + P P+ + P P S A V + + +

Sbjct: 585 PTLVASPPPVKSSPPAPVASPPPVKSSPPPTPVASPPPPAPVASSPPPMKSSPPPTP 644

Query: 167 AQPNGPEAEPRPPQSPASTASFIFSKGSRKLQLERPVSPETQADLQRLVAELRSISEQ 226  
PPE P PP PA + + ++ PE L+ +

Sbjct: 645 VSSPPPEKSP-PPPPAKSTPPPEEYTPPTS VKSSPPPEKSLP-PPTLIPSPPPQEK 702

Query: 227 RPPQAPKKSAPKPP-PVARKPSVGVPPASPSPYPRAEPLTAPP 268  
PP P K P +P P K V PP S P P+++PP

Sbjct: 703 TTPSTPSKPPSSPEKSPPEKSPVSSPPQTPKSSPPAPVSSPP 745

Score = 177 (26.6 bits), Expect = 2.6e-10, P = 2.6e-10  
Identities = 78/264 (29%), Positives = 105/264 (39%)

Query: 5 PPPEEAEFFSVASPEPAGP---SGSELVSSPAASSSATALQIQPGSP--DPPAP-- 56  
PPP +P+PA P S PE+V P+ + T I PP P PPP P

Sbjct: 850 PPAPVSSPPLTKPASPPAHVSSPPEVK-PSTPPAPTIV--ISPPSEPKSSPPPTV 906

Query: 57 -PAPAPASSAPGHVAKLPQKEPVGCSKGGPPREDVGAPLVTPSLQMVRLRSVGA 115  
P P SS P + P P PP V +P P++ V +P

Sbjct: 907 LPPPIVKSSPPAMVSSPPMTPKS-----SPPVVVSSP--PPTVKSSPPAPVSSPPAT 959

Query: 116 PTPALGPSAPQKPLRRALSGRASVPAPSSGLHAAVRLKACSLAASEGLSSAQPNPPEA 175  
P + P+ P ++SP P P S A + S +SS P PPE

Sbjct: 960 PKSSPPAPVNLPPPEV---KSSPPPTPVSSPPAPK---SSPPAPMSSP-P--PPEV 1009

Query: 176 EPRPPQSPASTASFIFSKGSRKLQLERPVSPETQADLQRLVAELRSISEQRPQAPKKS 235  
+ PP +P S+ + P P ++ V+ + PP AP S

Sbjct: 1010 KSPPPAPVSSPPPVKSSPPAPVSSP-PPPVKSSPPAPVSSPPPVKSSPPAPISS 1068

Query: 236 PKAPPPVARKPS---VGVPASPSPYPRAEPLTAPP 268  
P PPPV P V PPP S P P+++PP

Sbjct: 1069 P--PPPVKSSPPAPVSSPPPVKSSPPAPVSSPP 1102

Score = 177 (26.6 bits), Expect = 2.6e-10, P = 2.6e-10  
Identities = 82/267 (30%), Positives = 110/267 (41%)

Query: 17 PEPAG-PSGSELVSSPAASS---SSATALQIQPGSPDPPAP---PAPAPASSAPGHV 69  
P P G P SP + PAAS+ S T + P +P P P P P P +P

Sbjct: 410 PTPGGGPPSSP-VPGKPAASAPMPSPTTPDVSPPELPEPSVPAPAPMPTPHSPPAD 468

Query: 70 AKLPQKEPV-GCSKGGPPREDVGAPLVTPSLQMVRLRSVGA 128  
+P PV G S P V P + +V+L AP G+P P + ++P P

Sbjct: 469 DYVPPTTPVPGKSPATSPSQVQPPAASTPPSLVKLSPPQAPVGSPPPVKTTSPAP 528

Query: 129 LRRALSGRASVPAPSSGLHAAVRLKACSLAASEGLSSAQPNPPEAEPRPPQSPASTAS 188  
+ G SP P P S + +K+ A G + P PPE P PP AS

Sbjct: 529 I-----GSPSP-PPPVSVSSPPPVKSSPPAPVG---SPP--PPEKSPPPAPVASPP 577

Query: 189 FIFSKGSRKLQLERPV---SPETQADLQRLVAELRS-----ISEQRPQA-----PK 233  
+ S L P SP A + + ++S ++ PP P

Sbjct: 578 PVKSSPPPTLVASPPPVKSSPPAPVA-SPPPVKSSPPPTPVASPPPPAPVASSPPPM 636

Query: 234 KSPKAPPPVARKP---SVGVPPASPSPYPRAEPLTAPPTN 270  
KSP P PV+ P PPP + S P E PPT+

Sbjct: 637 KSPPTPVSSPPPEKSPPPPPAKSTPPPEEYTPPTS 676

Score = 170 (25.5 bits), Expect = 1.6e-09, P = 1.6e-09  
Identities = 78/279 (27%), Positives = 108/279 (38%)

Query: 5 PPPEEAFVSASPEPAGSGSPSELVSSPAASSSSATALQIQPGSPDPPAPAPAPASS 64  
 PP S S + P +P + P SS A+ PP +P +PP P SS  
 Sbjct: 883 PPAPTVISPPSEPKSSPPPTVSLPPPIVKSSPPAPMVSSPMTPKS--SPP-PVVVSS 939

Query: 65 APGHVAKLPQKEPVGCSKGGPPREDVGAPLVTPSLLQMVRLRSVGAPG--GAPTALGP 122  
 P V P PV PP +P P L ++S P +P PA  
 Sbjct: 940 PPPTVKSSPPAPVS-----SPPATPKSSPPAPVNLPPPEVKSSPPPTPVSSPPAPKS 994

Query: 123 SAPQKPLRRALSG--RASFPVPAPSSGLHAAVRLKACSLAASEGLSSAQNGPPEAEPRFP 180  
 S P P+ ++ P PAP S V+ S +SS P PP + PP  
 Sbjct: 995 SPPAPMSSPPPEVKSSPPAPVSSPPPVK----SPPAPVSS--P--PPPVKSPP 1046

Query: 181 QSPASTASFIFSKGSRKLQLERPVSPETQADLQRLNVAELRSISEQRPPQAPKSPKAPP 240  
 +P S+ + P P ++ V+ + PP AP SP PP  
 Sbjct: 1047 PAPVSSPPPVKSPPPPAPISSP-PPPVKSPPAPVSSPPPVKSPPPPAPVSSP--PP 1103

Query: 241 PVARKPS--VGVPASP--PSYPRAEPLTAPPTNGLPHTQDRTKREL 283  
 P+ P V PPPA PS P P+++PP P + ++ L  
 Sbjct: 1104 PPKSPPPPAPVSSPPAPVKPPLPPAPVSSPPPVTPAPPKKEEQL 1152

Score = 169 (25.4 bits), Expect = 2.1e-09, P = 2.1e-09  
 Identities = 75/266 (28%), Positives = 104/266 (39%)

Query: 3 DFPPEEAFVSASPEPAGSGSPSELVSSPAASSSSATALQIQPP----GSPDPP---PA 55  
 D+ PP V P S SP+ V PAAS+ + +++ PP GSP PP +  
 Sbjct: 469 DIVPPTTP---VPGKSPATSPSPQ-VQFFAASPPPSLVKLSPPQAPVGSPPPVKTT 524

Query: 56 PPAPAPASSAPGHVAKLPQKEPVGCSKGGPPREDVGAPLVTPSLLQMVRLRSVGAPGA 115  
 PPAP + S P V+ + PV PP VG+P P V+P  
 Sbjct: 525 PPAPIGSPSPFPVSVVSPPPPVKSP----PPAPVGSPP--PPPEKSPPPAPVASP--- 575

Query: 116 PTPALGPSAPQKPLRRALSGRASFPVPAPSSGLHAAVRLKACSLAASEGLSSAQNGPPEA 175  
 P P P P ++ P PAP + V+ S ++S P P +  
 Sbjct: 576 PPPVKSPPPTLVASPPPVKSPPPPAPVASPPPVK----SPPPTPVASPPPPAPVAS 631

Query: 176 EPRPQSPASTASFIFSKGSRKLQLERPVSPETQADLQRLNVAELRSISEQRPPQAPKKS 235  
 P P +SP K P P S+ PP+  
 Sbjct: 632 SPPMKSPPPPTPVSSPPPEKSP--PPPPAKSTPPEEYPTPTSVKSSPPPEKSLPP 689

Query: 236 PK---APPPVARK--PSVGVPASPSPYPR--EPLTAPP 268  
 P +PPP + PS PP+SP P EP+++PP  
 Sbjct: 690 PTLIPSPPPQEKPTPSTPSKPPSSPEKPSPPKEPVSSPP 729

Score = 168 (25.2 bits), Expect = 2.7e-09, P = 2.7e-09  
 Identities = 75/267 (28%), Positives = 102/267 (38%)

Query: 2 ADFPPEEAFVSASPE-PAGPSGSPSELVSSPAASSSSATALQIQPGSPDPP-PAPPAP 59  
 A PPP + ++ P+ P G P +SP A S + SP PP +PP P  
 Sbjct: 496 ASTPPP--SLVKLSPPQAPVGSPPPPVKTTSPPAPIGSPSPPPVSVVSPPPPVKSPPPP 553

Query: 60 APASSAPGHVAKLPQKEPVGCSKGGPPREDVGAPLVTPSLLQMVRLRSVGAPGAPGAPTA 119  
 AP S P P PV PP + P + S V+ AP +P P  
 Sbjct: 554 APVGSPPPEKSPPPAPVASPP--PPVKSPPPTLVASPPPPVKSPPPPAPVASPPPP 610

Query: 120 LGPSAPQKPLRRALSGRASFPVPAPSSGLHAAVRLKACSLAASEGLSSAQNGPPEAEPR 178  
 + P P+ + P PAP + ++ +S P PP A+  
 Sbjct: 611 VKSPPPPTPVA-----SPPPPAPVASSPPPMKSPPPPTVSSPPPEKSPPPPPAKST 664

Query: 179 PP--QSPASTASFIFSKGSRKLQLERPV--SPETQADLQRLNVAELRSISEQRPPQAPK 233  
 PP + P S S K L P SP Q S ++P +P  
 Sbjct: 665 PPPEEYPTPTTSVKSSPPPEK-SLPPPTLIPSPPPQEKPTPSTPSKPPSSPEK--SPP 721

Query: 234 KSPKAPPPVARKPSVGVPASPSPYPRAEPLTAPP 268  
 K P + PP K S PPPA S P P+++PP  
 Sbjct: 722 KEPVSSPPQTKSS---PPAPVSSPPPTPVSSPP 753

Score = 166 (24.9 bits), Expect = 4.6e-09, P = 4.6e-09  
 Identities = 81/268 (30%), Positives = 108/268 (40%)

Query: 5 PPPEEAF---FSVASPEPAGSGSPE-LVSSPAASSSS---ATALQIQPGSPDPP-- 54  
 PPPE++ VASP P S P LV+SP S A PP PPP  
 Sbjct: 560 PPPEKSPPPAPVASPPPVKSPPPTLVASPPPVKSPPPPAPVASPPPVKSPPPPPT 619

Query: 55 --APPAPAPASSAPGHVAKLPQKEPVGCSKGGPPREDVGAPLVTPSLLQMVRLRS 108  
 +PP PAP +S+P + P PV K PP P ++S  
 Sbjct: 620 VASPPPPAPVASSPPPMKSPPPPTPVSSPPPEKSPPPPPAKSTPPEEYPTPTTSVKS 679

Query: 109 VGAPGGA-PTPALGPSAPQKPLRRALSGRASFPVPAPSSGLHAAVRLKACSLAASEGLSSA 167  
 P + P P L PS P P + + ++P PSS + + S SS  
 Sbjct: 680 SPPPEKSLPPPTLIPSP--PQEKPTPSTPSKPPSSPEKSPPPKEPVSSPPQTKSSP 736

Query: 168 QPNGPPEAEPRPPQSPASTASFIFSKGSRKLQLERPVS PETQADLQRNLVAELRSISEQR 227  
 P P P SP + A + S S K P + P + + + +  
 Sbjct: 737 PPAVSSPPPTPVSSPPALAP-VSSPPSVKSS--PPAPLSSPPAPQVKSSPPPVQVSS 793

Query: 228 PPQAPKSPKAPPEVARKPSVGVPPSPSYPRAEPLTAPP 268  
 PP APK SP P+A P V PP + P PL++PP  
 Sbjct: 794 PPPAPKSSP---PLA--P-VSSPPQVEKTSPPAPLSSPP 827

Score = 165 (24.8 bits), Expect = 6.0e-09, P = 6.0e-09  
 Identities = 79/264 (29%), Positives = 105/264 (39%)

Query: 5 PPPEEAFSVASPEPAG-PSGSP--ELVSSPAASSSSATALQIQPPGSPDPPPPAPPAPA 60  
 PPP + + + P P G PS P +VS P S P GSP PP +PP PA  
 Sbjct: 517 PPPVK---TTSPPAPIGSPSPPPVSVSPPPPVKSPPPPA---PVGSPPPPEKSPPPPA 570

Query: 61 PASSAPGHVAKLPQKEPVGCSKG---GGPPREDVGAP---LVTPSLLQMVRLRSVGAPGG 114  
 P+S P V P V PP V +P + +P V AP  
 Sbjct: 571 PVASPPPPVKSPPPPTLVASPPPPVKSPPPAPVASPPPPVKSPPPPTPVASPPPPAPVA 630

Query: 115 APTPALGPSAPQKPLRRALSGRASVPVAP---SSGLHAAVRLKACSLAASEGLSSAQPNG 171  
 + P + P P+ SP P P S+ S+ +S + P  
 Sbjct: 631 SSPPPMKSPPPPTPVSSPPPEKSPPPPPAKSTPPPEEYPTPTSVKSSPPPEKSLP-- 688

Query: 172 PPEAEPRPPQSPASTASFIFSKGSRKLQLERPVS PETQADLQRNLVAELRSISEQRPPQA 231  
 PP P PP T SK P SPE + + V+ + PP A  
 Sbjct: 689 PTLIPSPPPQEKPTFPSTPSKP-----PSSPEKPS-PKEPVSSPPQTPKSSPPPA 739

Query: 232 PKKSPKAPPPVARKPSVGV--PPPASPSYPRAEPLTAPP 268  
 P SP P PV+ P++ PP+ S P PL++PP  
 Sbjct: 740 PVSSPP-PTPVSSPPALAPVSSPPSVKSSPPAPLSSPP 777

Score = 162 (24.3 bits), Expect = 1.3e-08, P = 1.3e-08  
 Identities = 76/272 (27%), Positives = 99/272 (36%)

Query: 2 ADFPPPEEAFSVASPEPAG-PSGSP-ELVSSPAASSSSATALQIQPPGSPDPPPPAPPAPA 60  
 A P P SPEP PS P P + S A PP P P +PPA +  
 Sbjct: 427 ASAPMPSPHTPPDVSPLEPEPSVPAPAPMPMPTPHSPPADDDYVPTPPVPGKSPPTS 486

Query: 61 PASSAPGHVAKLPQKEPVGCSKGGPPREDVGAPLVTPSLLQMVRLRSVGAPGGAPT-- 118  
 P+ A P V S PP+ VG+P P V+ S AP G+P+P  
 Sbjct: 487 PSPQVQPPAATPPPSLVKLS---PPQAPVGSF--PPF---VKTSPPAPIGSPSPPP 536

Query: 119 ---ALGPSAPQK-PLRRALSGRASVPVAPSSGLHAAVRLKACSLAASEGLSSAQPNGPPE 174  
 + P P K P A G SP P S A S + + PP  
 Sbjct: 537 PVSVPSPPPVKSPPPAPVG--SPPPEKSPPPAPVASPPPPVKSPPPPTLVASPPPP 594

Query: 175 AEPRPPQSPASTASFIFSKGSRKLQLERPVS PETQADLQRNLVAELRSISEQRPPQAPKK 234  
 + PP +P + + + P P A + + PP P+K  
 Sbjct: 595 VKSPPPAPVASPPPPVKSPPPPTPVASPPPPAPVASSPPPMKSPPPPTPVSSPPP-PEK 653

Query: 235 SPKAPPPVARKPSVGVPPSPSYPRAEPLTAPPTNGLP 273  
 SP PPP P PP P+ P + + PP LP  
 Sbjct: 654 SPPPPPPAKSTP---PPEEYPTPTSVKSSPPPEKSLP 688

Score = 159 (23.9 bits), Expect = 2.8e-08, P = 2.8e-08  
 Identities = 77/264 (29%), Positives = 103/264 (39%)

Query: 5 PPPEEAFSVASPEPAGPSGSP-ELVSSPAASSSSATALQIQPPGSP--DPPAP---PAP 59  
 PPP V+SP P P SP P SS ++ PP +P PP P P P  
 Sbjct: 916 PPPA---MVSSP-PMTPKSSP---PVVSSPPPTVKSSPPAPVSSPPATPKSSPP 966

Query: 60 APASSAPGHVAKLPQKEPVGCSKGGPPREDVGAPLVTPSLLQMVRLRSVGAPGGAPT 119  
 AP + P V P PV S P AP+ +P + V+ AP +P P  
 Sbjct: 967 APVNLPPPEVKSSPPPTPV-SPPFAPKSSPPAPMSSPPPE-VKSPPPAPVSSPPPP 1024

Query: 120 LGPSAPQKPLRRALSG-RASVPVAPSSGLHAAVRLKACSLAASEG---LSSAQPNGPPEA 175  
 + P P+ ++ P PAP S V+ S + S P P +  
 Sbjct: 1025 VKSPPPAPVSSPPPPVKSPPPAPVSSPPPPVKSPPPAPISSPPPPVKSPPPAPVSS 1084

Query: 176 EPRPPQSPASTASFIFSKGSRKLQLERPVS PETQADLQRNLVAELRSISEQRPPQAPKKS 235  
 P P +SP A S ++ P P A + A ++ S PP AP S  
 Sbjct: 1085 PPPPVKSPPPAPV---SSPPPIKSPPP---APVSSPPAPVKPS--LPPAPVSS 1135

Query: 236 PK--APPPVARKPSVGVPPPA-SPSYPRAEPLTAPP 268  
 P P +K +PPPA S P + PP  
 Sbjct: 1136 PPPVVTAPPKKEQSLPPFAESQPPSFNDIILPP 1171

Score = 143 (21.5 bits), Expect = 1.8e-06, P = 1.8e-06  
 Identities = 59/179 (32%), Positives = 77/179 (43%)

Query: 3 DFPPEEAFSSVASPEP-AGPSGSELVSSPAASSSSATA-LQIOPPGSP--DPPP---A 55  
 + PPPE S P P + P +P+ PA SS ++ PP +P PPP +  
 Sbjet: 970 NLPPPEVK--SSPPTPVSSPPAPKSSPPAPMSSPPPEVKSPPPAPVSSPPPPVK 1027

Query: 56 PPAPAPASSAPGHVAKLPQKEPVGCSKGGPPREDVGAPLVTPLQLQMVRLRSVGAPGGA 115  
 PP PAP SS P V P PV PP + P S V+ AP +  
 Sbjet: 1028 PPPAPVSSPPPPVKSPPPAPVSSPP---PPVKSPPPPAPISSPPPPVKSPPPAPVSS 1084

Query: 116 PTPALGPSAPQKPLRRALSG-RASVPAPSSGLHAAVRLKACSLAASEGLSSAQNGPPE 174  
 P P + P P+ ++ P PAP S A +K SL +SS P PP  
 Sbjet: 1085 PPPPVKSPPPPAPVSSPPPIKSPPPAPVSSPPAP-VKPSLPPAPVSS--P--PPV 1139

Query: 175 AEP RPQ 181  
 P PP+  
 Sbjet: 1140 VTPAPPK 1146

Score = 133 (20.0 bits), Expect = 2.3e-05, P = 2.3e-05  
 Identities = 50/132 (37%), Positives = 59/132 (44%)

Query: 1 MADFPPEEAFSSVASPEPAGP-SGSELVSSP---AASSSSATALQIOPPGSP--DPPP 54  
 M+ PPPE V SP P P S P V SP A SS ++ PP +P PPP  
 Sbjet: 1001 MSSPPPE---VKSPPPAPVSSPPPPVKSPPPAPVSSPPPPVKSPPPAPVSSPP 1055

Query: 55 ---APPAPAPASSAPGHVAKLPQKEPVGCSK---GGPPREDVGAPLVTPLQLQMVRLRS 108  
 +PP PAP SS P V P PV PP V +P P +  
 Sbjet: 1056 PVKSPPPPAPISSPPPPVKSPPPAPVSSPPPPVKSPPPAPVSSP--PPPIKSPPPAP 1113

Query: 109 VGAPGGAPT--PALGPSAP 125  
 V +P AP P+L P AP  
 Sbjet: 1114 VSSPPAPVKPSLPPAP 1132

Score = 110 (16.5 bits), Expect = 8.0e-03, P = 8.0e-03  
 Identities = 41/121 (33%), Positives = 49/121 (40%)

Query: 5 PPPEEAFSS---VASPEPAGP-SGSELVSSP---AASSSSATALQIOPPGSP--DPPP 54  
 PPP S V SP P P S P V SP A SS ++ PP +P PPP  
 Sbjet: 1060 PPPPAPISSPPPPVKSPPPAPVSSPPPPVKSPPPAPVSSPPPIKSPPPAPVSSPP 1119

Query: 55 AP-----PAPAPASSAPGHVAKLPQKEPVGCSKGGPPREDVGAPLVTPLQLQMVRLRS 108  
 AP P PAP SS P V P K+ + PP E P +L +  
 Sbjet: 1120 APVKPSLPPAPVSSPPPVVTPAPPKKE---EQSLPPPAESQPPSFNDIILPIMANK 1176

Query: 109 VGAP 112  
 +P  
 Sbjet: 1177 YASP 1180

Score = 108 (16.2 bits), Expect = 1.3e-02, P = 1.3e-02  
 Identities = 46/155 (29%), Positives = 67/155 (43%)

Query: 114 GAPTPALGPSAPQKPLRRALSGRASVPAPSSGLHAAVR-LKACS-LAASEGLSSAQNG 171  
 G PTP GP + P + A S +P+P+ + L S + A + P+  
 Sbjet: 408 GYPTPGGGPPSSPVGKPAAS---APMPSHTPPDVSPLEPLEPSVPVAPAPMPMPTPHS 464

Query: 172 PPEAEPRPPQSPASTASFISKGSRLQLERPVSFETQ---ADLQRNLVAELRSISEQR 227  
 PP + PP P S + S ++Q +P + Q + + +  
 Sbjet: 465 PPADDYVPPTPPVPGKSPATSPSPQVQPPAASTPPPSLVKLSPPQAPVGSPPPPVKTTS 524

Query: 228 PPQAPKKSPPKAPPPVARKPSVGVPPPASPSYPRAEPLTAPP 268  
 PP AP SP PPPV SV PPP S P P+ +PP  
 Sbjet: 525 PP-APIGSPSPPPV---SVVSPPPVKSPPPAPVGSPP 560

Pedant information for DKF2phmcfl\_1c23, frame 1

Report for DKF2phmcfl\_1c23.1

[LENGTH] 311  
 [MW] 31534.58  
 [pI] 9.48  
 [KW] All\_Alpha  
 [KW] LOW\_COMPLEXITY 38.59 %

SEQ MADFPPEEAFSSVASPEPAGPSGSELVSSPAASSSSATALQIOPPGSPDPPAPAPAPA  
 SEG .....XX  
 PRD ccc

SEQ PASSAPGHVAKLPQKEPVGCSKGGPPREDVGAPLVTPLQLQMVRLRSVGAPGGAPTAL  
 SEG xxxxxx.....XX



(No Prosite data available for DKFZphmcf1\_1c23.1)  
(No Pfam data available for DKFZphmcf1\_1c23.1)

DKF2phmcf1\_1e15

group: transmembrane protein

DKF2phmcf1\_1e15 encodes a novel 454 amino acid protein with similarity to *C. elegans* proteins and transporter proteins.

The novel protein is similar to the PTR2 family of proton/oligopeptide symporter proteins and the D-xylose-proton symporter. Thus, the protein is a transporter of a so far unknown compound.

The new protein can find application as a new transporter in eukaryotic cells, e.g. in drug transport into cells.

similarity to D-XYLOSE TRANSPORTER  
membrane regions: 9

complete cDNA, complete cds, EST hits  
matches cDNA encoding cell growth inhibiting factor (E12646)

Sequenced by DKF2

Locus: unknown

Insert length: 1957 bp

Poly A stretch at pos. 1947, polyadenylation signal at pos. 1929

```

1  GGTGCAGCGC  CCGGCTGAG  CGACAGCAAG  TGCAGCGGGC  TCCTACCCCG
51  GGTGAGGGGT  GGCTCCGCG  TGGGATCGTG  CCCTCTTCAG  CCCGCTCCTG
101  TCCCGACAT  CACGTGTATT  CCGCAGCTCC  CCTCCGCGCT  GTGTGCTTAC
151  TGAGACGGGG  AGGCGTGACA  GGGCCGGGGT  CCCTTCTCAG  TGGTGCTCTG
201  TGCTTCAGGG  CAAGCTCCCC  GTCTCCGGGG  GCACTTCCTT  CGCTGTGTGT
251  CGGTCCATCC  TCCTTTCTCC  AGCTCTCTCC  CCTCGCAGGT  GGGATCGTGG
301  GTGGGACCGG  AGCGCGGGCG  GCGCGGGCCC  CCGGGGACCA  TGGCGGGGTC
351  CGACACCGCG  CCCTTCTCA  GCCAGGCGGA  TGACCCGGAC  GACGGGCGAG
401  TGCTTGGCAC  CCGGGGGTTG  CCAGGGTCCA  CGGGGAACCC  GAAGTCCGAG
451  GAGCCCGAGG  TCCCGGACCA  GGAGGGGCTG  CAGCGCATCA  CCGGCTGTTC
501  TCCCGGCCGT  TCGGCTCTCA  TAGTGGCGGT  GCTGTGCTAC  ATCAATCTCC
551  TGAACATCAT  GGACCGCTTC  ACCGTGGCTG  TGTTCATCTC  CAGTTACATG
601  GTGTTGGCAC  CTGTGTTTGG  CTACCTGGGT  GACAGGTACA  ATCGGAAGTA
651  TCTCATGTGC  GGGGGCATTG  CCTTCTGGTC  CTTGGTGACA  CTGGGGTCTA
701  CCTTCATCCC  GGGAGAGCAT  TTCTGGCTGC  TCCTCCTGAC  CCGGGGCTTG
751  GTGGGGGTGC  GGGAGGCGAG  TTATTCCACC  ATCGCGCCCA  CTCTCATTGC
801  CGACCTCTTT  GTGGCCGACC  AGCGGAGCCG  GATGCTCAGC  ATCTTCTACT
851  TTGCATTTC  GGTGGGCGAG  GGTCTGGGCT  ACATTGCAGG  CTCCAAAGTG
901  AAGGATATGG  CTGGAGACTG  GCACTGGGCT  CTGAGGGTGA  CACCGGGTCT
951  AGCAGTGTG  GCCGTTCTGC  TGCTGTCTCT  GGTAGTGGCG  GAGCCGCCAA
1001  GGGGAGCCGT  GGAGCGCCAC  TCAGATTTCG  CACCCCTGAA  CCCCACCTCG
1051  TGGTGGGCAG  ATCTGAGGGC  TCTGGCAAGA  AATCTCATCT  TTGGACTCAT
1101  CACCTGCCTG  ACCGGAGTCC  TGGGTGTGGG  CCTGGGTGTG  GAGATCAGCC
1151  GCCGGCTCCG  CCACTCCAAC  CCGGGGCTTG  ATCCCTGGT  CTGTGCCACT
1201  GGCTCTCTGG  GCTCTGCACC  CTTCTCTTTC  CTGTCCCTTG  CCTGCGCCCG
1251  TGGTAGCATC  GTGGCCACTT  ATATTTTCAT  CTTTATTGGA  GAGACCCCTC
1301  TGTCATGAA  CTGGGCCATC  GTGGCCGACA  TTCTGCTGTA  CGTGGTGATC
1351  CCTACCCGAC  GCTCCACCGC  CGAGGCCCTT  CAGATCGTGC  TGTCCCACCT
1401  GCTGGGTGAT  GCTGGGAGCC  CCTACCTCAT  TGGCTGATC  TCTGACCGCC
1451  TGCGCCGGAA  CTGGCCCCC  TCCTTCTTGT  CCGAGTTCCG  GGCTCTGCAG
1501  TTCTCGTCA  TGCTCTGCGC  GTTTGTTGGG  GCACTGGGCG  GCGCAGCCTT
1551  CCTGGGCACC  GCCATCTTCA  TTGAGGCCGA  CCGCCGGCGG  GCACAGCTGC
1601  ACGTGCAGGG  CTTGCTGCAC  GAAGCAGGGT  CCACAGACGA  CCGGATTGTG
1651  GTGCCCCAGC  GGGGCCGCTC  CACCCGCTGT  CCGTGGCCA  GTGTGCTCAT
1701  CTGAGAGGCT  GCCGCTCACC  TACCTGCACA  TCTGCCACAG  CTGGCCCTGG
1751  GCCCACCCCA  CGAAGGGCCT  GGGCCTAACC  CTTGGGCTTG  GCCCAGCTTC
1801  CAGAGGGACC  CTGGGCCGTG  TGCCAGCTCC  CAGACACTAC  ATGGGTAGCT
1851  CAGGGGAGGA  GGTGGGGGTC  CAGGAGGGGG  ATCCCTCTCC  ACAGGGGCGA
1901  CCCCAGGGGC  TCGGTGCTAT  TTGTAACGGA  ATAAATTTG  TAGCCAGAAA
1951  AAAAAAA

```

## BLAST Results

Entry E12646 from database EMBL:  
cDNA encoding cell growth inhibiting factor.  
Score = 3046, P = 2.2e-131, identities = 640/659

## Medline entries

No Medline entry

## Peptide information for frame 1

ORF from 340 bp to 1701 bp; peptide length: 454  
 Category: similarity to known protein

```

1 MAGSDTAPFL SQADDDDDGP VPGTPGLPGS TGNPKSEEP E VPDQEGLRQI
51 TGLSPGRSAL IVAVLCYINL LNYMDRFTVA VFISYMYLA PVFGYLGDRY
101 NRKYLKCGGI AFWSLVTLGS SFIPGHEFWL LLLTRGLVGV GEASYSTIAP
151 TLIADLFVAD QRSRLSIFY FAIPVGSGLG YIAGSKVKDM AGDWHWALRV
201 TPGLGVVAVL LFLVVRREPP RGAVERHSDI PPLNPTSWWA DRLALARNLI
251 FGLITCLTGV LGVGLGVEIS RRLRHSNPRA DPLVCATGLL GSAPFLFLSL
301 ACARGSIVAT YIFIFIGETL LSMNWAIVAD ILLYVVIPT RSTAEAFQIV
351 LSHLLGDAGS PYLIGLISDR LRRNWPPSFL SEFRALQFSL MLCATFVGALG
401 GAAFLGTAFI IEADRRRAQL HVQGLLHEAG STDDRIVVPQ RGRSTRVPVA
451 SVLI

```

## BLASTP hits

No BLASTP hits available

## Alert BLASTP hits for DKFZphmcf1\_1e15, frame 1

TREMBL:CEC13C4\_1 gene: "C13C4.5"; Caenorhabditis elegans cosmid C13C4,  
 N = 3, Score = 441, P = 5.2e-76

TREMBL:CEC39E9\_10 gene: "C39E9.10"; Caenorhabditis elegans cosmid  
 C39E9, N = 2, Score = 449, P = 8.2e-69

TREMBL:CEF09A5\_1 gene: "F09A5.1"; Caenorhabditis elegans cosmid F09A5,  
 N = 3, Score = 413, P = 9.1e-60

TREMBL:ATF6H11\_18 gene: "F6H11.180"; product: "predicted protein";  
 Arabidopsis thaliana DNA chromosome 5, BAC clone F6H11 (ESSAII  
 project), N = 3, Score = 193, P = 2.5e-24

SWISSPROT:XYLT LACBR D-XYLOSE-PROTON SYMPORT (D-XYLOSE TRANSPORTER)., N  
 = 1, Score = 180, P = 7.9e-11

>TREMBL:CEC39E9\_10 gene: "C39E9.10"; Caenorhabditis elegans cosmid C39E9  
 Length = 488

## HSPs:

Score = 449 (67.4 bits), Expect = 8.2e-69, Sum P(2) = 8.2e-69  
 Identities = 88/204 (43%), Positives = 125/204 (61%)

Query: 58 SALIVAVLCYINLLNYMDRFTVAVFISSYMYLAPVFGYLGDRYNRKYLMCGGIAFWSLVT 117  
 + ++ V Y N + + + VF+ S+MV+PV GYLGD+NRK+M G+ W

Sbjct: 29 AGVLTQVQTYYNISDSLGLLIQTVFLISEFMVSPVCGYLGDRFNKRWIMIGVGIWLGAV 88

Query: 118 LGSSFIPEGHEFWL LLLTRGLVGVGEASYSTIAPTILADLFVADQRSRLSIFYFAIPVGS 177  
 LGSSF+P HFWL L+ R VG+GEASYS +AP+LI+D+F +RS + IFYFAIPVGS

Sbjct: 89 LGSSFVPANHFVFLVLRSFVGIGEASYSNVAPSLISDMFNGQKRSTVFMIFYFAIPVGS 148

Query: 178 GLGYIAGSKVKDMAGDWHWALRVTPGLGVVAVLLFLVVRREPPRGAVR----HSDLPL 233  
 GLG+I GS V + G W W +RV+ G++ ++ L L EP RGA ++ D+

Sbjct: 149 GLGFIVGNSVATLTGHVQWQIRVSAIAGLIVMIALVLFTYEPERGAADKAMGESKDVVVT 208

Query: 234 NPTSWWADRLALARNLIFGLITCLTG 259

T++ DL L + L+ C G

Sbjct: 209 TNTTYLEDLVILLKTPT--LVACTWG 232

Score = 267 (40.1 bits), Expect = 8.2e-69, Sum P(2) = 8.2e-69  
 Identities = 74/212 (34%), Positives = 113/212 (53%)

Query: 249 LIFGLITCLTGLVGLGVEISRRL-----RHSNPRADPLVCATGLLGSAPFLFLSL 300  
 L FG IT G++GV G +S+ L R RA PLV G L +APFL + +

Sbjct: 277 LYFGAITTAGGLIGVIFGSMLSKWLVAGWGPFRRLQTDRAQPLVAGGALLAAPFLIGM 336

Query: 301 ACARGSIVATYIFIFIGETLLSMNWAIVADILLYVVIPTRRSTAEAFQIVLSHLLGDAGS 360

Pedant information for DKFZphmcf1\_1e15, frame 1

```
[LENGTH]           454
[MW]                49013.35
[pI]                7.66
[HOMOL]             TREMBL:CEC13C4_1 gene: "C13C4.5"; Caenorhabditis elegans cosmid C13C4 2e-51

[BLOCKS]            BL01022D
[PROSITE]            MYRISTYL             11
[PROSITE]            CAMP_PHOSPHO_SITE     1
[PROSITE]            CK2_PHOSPHO_SITE      3
[PROSITE]            PROKAR_LIPOPROTEIN    1
[PROSITE]            GLYCOSAMINOGLYCAN     1
[PROSITE]            PKC_PHOSPHO_SITE      4
[PROSITE]            TRANSMEMBRANE         8
[KW]                 LOW_COMPLEXITY        15.42 %
```

556

SEG .....  
PRD hhhhhhhcccccccccccccccccccccc  
MEM MMMMMMMMMMMMMMMMMMMMMMMMMMM.....

## Prosite for DKFZphmcf1\_1e15.1

PS00002	177->181	GLYCOSAMINOGLYCAN	PDOC00002
PS00004	340->344	CAMP_PHOSPHO_SITE	PDOC00004
PS00005	270->273	PKC_PHOSPHO_SITE	PDOC00005
PS00005	339->342	PKC_PHOSPHO_SITE	PDOC00005
PS00005	368->371	PKC_PHOSPHO_SITE	PDOC00005
PS00005	444->447	PKC_PHOSPHO_SITE	PDOC00005
PS00006	11->15	CK2_PHOSPHO_SITE	PDOC00006
PS00006	342->346	CK2_PHOSPHO_SITE	PDOC00006
PS00006	431->435	CK2_PHOSPHO_SITE	PDOC00006
PS00008	26->32	MYRISTYL	PDOC00008
PS00008	32->38	MYRISTYL	PDOC00008
PS00008	52->58	MYRISTYL	PDOC00008
PS00008	139->145	MYRISTYL	PDOC00008
PS00008	176->182	MYRISTYL	PDOC00008
PS00008	252->258	MYRISTYL	PDOC00008
PS00008	262->268	MYRISTYL	PDOC00008
PS00008	266->272	MYRISTYL	PDOC00008
PS00008	288->294	MYRISTYL	PDOC00008
PS00008	305->311	MYRISTYL	PDOC00008
PS00008	397->403	MYRISTYL	PDOC00008
PS00013	292->303	PROKAR_LIPOPROTEIN	PDOC00013

(No Pfam data available for DKFZphmcf1\_1e15.1)

DKFZphmcfl\_lg13

-----

group: mammary carcinoma derived

DKFZphmcfl\_lg13 encodes a novel 573 amino acid protein with very weak similarity to the human KIAA0543 protein and Musca domestica hermes transposase.

No informative BLAST results; No predictive prosite, pfam or SCOP motif.

The new protein can find application in studying the expression profile of mammary carcinoma-specific genes.

similarity to KIAA0766

complete cDNA, complete cds, few EST hits  
on genomic level encoded by AC005020, no splicing, genomic?

Sequenced by DKFZ

Locus: unknown

Insert length: 2210 bp

Poly A stretch at pos. 2200, polyadenylation signal at pos. 2176

```

1  GAAACCTGAT CTCATAAAAC CTAGGTCACA AAGGACAGCC CTGCAAAACA
51  GACCCATATT GGATCAAGTG AGCCAGTTCC TGGAACTGA ATAATGACTC
101 CTGAATCAAG GGATACTACA GATTTGTCTC CAGGGGCTAC CCAGGAGATG
151 GAAGGCATCG TGATAGTGAA GGTGGAGGAG GAAGATGAAG AAGACCATT
201 TCAAAAGGAA AGAAACAAAG TAGAGTCATC GCCACAAGTT CTCAGTCGCT
251 CTCAACTAT GAATGAGAGA GCCTTATTGT CATCGTATTT AGTTGCAAT
301 AGAGTGGCAA AAGAGAAAAT GGCTCACACA CGCGCTGAAA AAATTATCCT
351 TCCAGCATGT ATGGACATGG TACGGACAAT TTTTGATGAC AAATCAGCTG
401 ATAAACTAAG AACTATACCT CTTAGTGATA ATACAATATC TCGTCGAATC
451 TGTACGATTG CAAAACATT GGAAGCAATG CTTATTACAC GGCTGCAGTC
501 CGGTATAGAC TTGCAATCC AACTCGATGA GAGCACTGAT ATTGCAAGTT
551 GTCCCACTCT CTGCTTTTAT GTCAGATATG TGTGGCAAGA TGATTTTGTA
601 GAGGATCTCT TATGTTGTTT AAATTTAAAT TCACATATAA CTGGATTAGA
651 TTTATTTACT GAATTAGAAA ACTGCCTTCT TGGTCAGTAT AAATTAAGCT
701 GGAACATTG TAAAGGAATT TCAAGTCATG GAACAGCAAA TATGACCGGA
751 AAACACAGCA GACTTACTGA AAAATTGTTA GAAGCAACCC ACAACATGCG
801 TGTTTTGAAT CACTGTTTGA TTCAATCGAG AGCTTTGGTA TCCAAAGAAA
851 TTTACCAAG TCTGATGGAT GTATTGAAA ATGCAGTGAA AACTGTTAAT
901 TTTATTAAG GAAGCTCACT GAATAGCCGA CTTCTCGAAA TATTTGTTTC
951 AGAGATTGGA GTGAACACA CCCACTTATT GTTTCATACA GAAGTTCGTT
1001 GGCCTTCTCA AGGAAAAGTA TTGAGCAGAG TATATGAAC CAGGAACGAG
1051 ATTTACATTT TTCTCGTTGA AAAGCAATCT CATTGGCAAA ATATTTTGGA
1101 AGACGACATT TGGGTAACAA AATTGGCATA TTTAAGTGAT ATTTTGGCA
1151 TTCTTAATGA ATTAAGCCTG AAAATGCAGG GGAAAAACAA TGATATATTT
1201 CAGTATCTTG AACATAITCT AGGATTCCAA AAGACGTTAT TATTGTGGCA
1251 AGCAAGACTT AAAAGTAACC GCCCTAGCTA CTATATGTTT CCAACATTAT
1301 TGCAACACAT CGAAGAGAAC ATTATTAATG AAGACTGCTT AAAAGAAATA
1351 AAATTAGAGA TATTGTTGCA TCTCACTTCT TTGCTCAAAA CTTTTAATTA
1401 TTAATTTCCG GAAGAGAAAT TTGAATCATT AAAGGAAAT ATTTGGATGA
1451 AAGATCCATT TGCTTTTCAA AACCCAGAAAT CAATAATTGA GTTAAACTTG
1501 GAGCCTGAAG AAGAGAAATGA ATTATTGCAG CTCAGTTTAT CATTACACT
1551 AAAGAATTAT TATAAGATAT TAAGTTTATC AGCATTTTGG ATTAAGATTA
1601 AAGATGACTT TCCACTGCTA AGTAGGAAGA GTATATTGCT GTTACTACCA
1651 TTCACAACCTA CATATTGTG TGAAGTAGGA TTTTCAATCT TGACACGGTT
1701 AAAACAAAG AAGAGAAATA GGCTCAATAG TGCACAGAT ATGCGGGTAG
1751 CATTATCTTC ATGTGTTTCT GACTGGAAGG AACTTATGAA CAGACAAGCA
1801 CACCATCAC ATTAATATCA AACTTTACAA AATTCTGTGT ATAGCCAGGT
1851 TGGTGGCTT ACGCCTGTAA TCCAGCAGT GGGAGACCGA GGTGGGCAGA
1901 TCACTTGAGT TCAAGACCAG CCTGGCCAAC ATGGTGAAAC CCCATCTCTA
1951 CTAATAATAG AAACCTTAGC CAGGCGTGGT GGCACATGCC TGCAGTCCCA
2001 GTTACTTGGG TGCTGAGGC AGGAGAACTC CTTAAACCA GAAAGGCAGG
2051 ATTGCAGTGA GCTGAGATAA TCCCAGTGCA TTCCAGCCTG GGCAACAGCG
2101 TGAGACTTCA TCTCAAAAAA AAAAATTGTT ATTTGTACTT TTAAGGGGAT
2151 TTTGAGTAT GTTGTAGTTA AACGTTAATA AAATTATATT TGTAATTAGG
2201 AAAAAAAAAA

```

## BLAST Results

-----

Entry AC005020 from database EMBL:  
Homo sapiens clone GS259H13; HTGS phase 1, 4 unordered pieces.  
Score = 9110, P = 0.0e+00, identities = 1822/1822

## Medline entries

No Medline entry

## Peptide information for frame 1

ORF from 94 bp to 1812 bp; peptide length: 573  
 Category: similarity to unknown protein

```

1 MTPESRDTTD LSPGGTQEME GIVIVKVEEE DEEDHFQKER NKVESSPQVL
51 SRSTTMNERA LLSSYLVAIR VAKEKMAHTA AEKIILPACM DMVRTIFDDK
101 SADKLRTIPL SDNTISRRIC TIAKHLEAML ITRLQSGIDF AIQLDESTDI
151 ASCPTLLVYV RYVWQDDFVE DLLCCLNLNS HITGLDLFTE LENCLLGQYK
201 LNWKKHKGIS SDGTANMTGK HSRLTEKLEL ATHNNVWNN CFHREALVS
251 KEISPSLMDV LKNAVKTVMF IKGSSLNSRL LEIFCSEIGV NHTHLFHTF
301 VRWLSQGVKL SRVYELRNEI YIFLVEKQSH LANIFEDDIW VTKLAYLSDI
351 FGILNELSLK MQGKNNDIFQ YLEHILGFQK TLLWQARLK SNRPSYMF
401 TLLQHIENI INEDCLKEIK LEILLHLTSL SOTFNYYFPE EKFSLEKENI
451 WMKDPFAFQN PESIIEINLE PEEENELLQL SSSFTLKNNY KILSLAFWI
501 KIKDDPFLS RKSILLLLPF TTYLCELGF SILTRLKTKK NRRLNSAPDM
551 RVALSSCVDP WKELMNRQAH PSH

```

## BLASTP hits

Entry AC004877\_3 from database TREMBLNEW:  
 gene: "WUGSC:H DJ0751H13.2"; product: "KIAA0543 protein"; Homo sapiens  
 PAC clone DJ0751H13 from 7q35-qter, complete sequence.  
 Score = 86, P = 4.4e-03, identities = 46/179, positives = 78/179

Entry MD36211\_1 from database TREMBL:  
 product: "Hermes transposase"; Musca domestica Hermes transposase  
 gene, complete cds.  
 Score = 105, P = 3.0e-02, identities = 101/465, positives = 202/465

## Alert BLASTP hits for DKFZphmcf1\_lg13, frame 1

TREMBL:AB018309.1 gene: "KIAA0766"; product: "KIAA0766 protein"; Homo  
 sapiens mRNA for KIAA0766 protein, complete cds., N = 1, Score = 300, P  
 = 1.1e-23

>TREMBL:AB018309.1 gene: "KIAA0766"; product: "KIAA0766 protein"; Homo  
 sapiens mRNA for KIAA0766 protein, complete cds.  
 Length = 607

## HSPs:

Score = 300 (45.0 bits), Expect = 1.1e-23, P = 1.1e-23  
 Identities = 120/485 (24%), Positives = 229/485 (47%)

```

Query:      89 CMD-MVRTIFDDKSADKLRTIPLSDNTISRRICTIAKHLEAMLITRLQSGIDFAIQLDES 147
           CM+ ++R + + L+ + LS + +RI +I ++L L R + +++ LD+
Sbjct:     124 CMEVLLREVLPEH-VSVLQGVLDSPDITRQILSIDRNLRLNQLFNARDFKAYSALDDQ 182

Query:     148 TDIASCPTLLVYVRYVWQD-DFVEDLLCCLNLNSHIT-GLDLFTELENCLLGQYKLNWKH 205
           +A LLV++R V + + EDLL +NL H + G + LE+ L L+ +
Sbjct:     183 AFVAYENYLLVFIRGVGPELEVQEDLLTIINLTHFVSVALMSAILES--LQTAGLSLQR 240

Query:     206 KGISSDGTANMTGKHSRLTEKLEATHNNVAVN--HC--FIHREALVSKEISPSLMDVL 261
           G+++ T M G++S L + E + WN H F+H E L S ++ + ++
Sbjct:     241 MVGLTTTHTLRMIGENSGLSYMREKAVSPNCWNVIHYSGLHLELLSSYDQVDN--QII 298

Query:     262 KNAVKTVMFVIGSSLSNRLLEIFCSEIGVNHHTLHFTFTEVR-WLSQGVLSRVYELRNEI 320
           + + IK + + +E H + + WL +GK L ++ LR E+
Sbjct:     299 NTISEWIVLIKTRGVRRPEFQTLTSESEHGERVNGRCNLNWLRRGKTLKLIFSLRKEM 358

Query:     321 YIFLVEKQSHLANIFEDDIWVTKLAYLSDFGILNELSLKMQGKNNDIFQYLEHILGFQK 380
           FLV + + + F D W+ +L DI L ELS +++ +HI F+
Sbjct:     359 EAFVLSVGATTVH-FSDQWLCDGFLVDIMEHLRELSEELRVSKVFAAAAFDHICTFEV 417

```

Score = 290 (43.5 bits), Expect = 1.5e-22, P = 1.5e-22  
Identities = 120/485 (24%), Positives = 228/485 (47%)

Pedant information for DKFZphmcf1\_lg13, frame 1

```

[LENGTH]      573
[MW]           66276.85
[pI]           5.82
[HOWOL]       TREMBL:AB018309.1 gene: "KIAA0766"; product: "KIAA0766 protein"; Homo sapiens
mRNA for KIAA0766 protein, complete cds. 1e-18
[PROSITE]     MYRISTYL      3
[PROSITE]     CK2_PHOSPHO_SITE      10
[PROSITE]     TYR_PHOSPHO_SITE      1
[PROSITE]     PKC_PHOSPHO_SITE      9
[PROSITE]     ASN_GLYCOSYLATION      2
[KW]           All_Alpha
[KW]           LOW_COMPLEXITY      8.90 %

SEQ      MTPESRDTTDLSPGGTQEMEGIVIVKVEEEDDEHDFQKERNKVESPQVLSRSTTMNERA
SEG      .....XXXXXXXXX.....
PRD      ccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc
        ccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc

SEQ      LLSSYLVAYRVAKEMAHTAAEKIILPACMDMVRTIFDDKSADKLRTIPLSDNTISRRIIC

```



[illegible]

Prosite for DKFZphmcf1\_lg13.1

PS00001	216->220	ASN GLYCOSYLATION	PDCC000001
PS00001	291->295	ASN GLYCOSYLATION	PDCC000001
PS00005	116->119	PKC PHOSPHO_SITE	PDCC000005
PS00005	218->221	PKC PHOSPHO_SITE	PDCC000005
PS00005	225->228	PKC PHOSPHO_SITE	PDCC000005
PS00005	358->361	PKC PHOSPHO_SITE	PDCC000005
PS00005	391->394	PKC PHOSPHO_SITE	PDCC000005
PS00005	445->448	PKC PHOSPHO_SITE	PDCC000005
PS00005	485->488	PKC PHOSPHO_SITE	PDCC000005
PS00005	510->513	PKC PHOSPHO_SITE	PDCC000005
PS00005	538->541	PKC PHOSPHO_SITE	PDCC000005
PS00006	55->59	CK2 PHOSPHO_SITE	PDCC000006
PS00006	79->83	CK2 PHOSPHO_SITE	PDCC000006
PS00006	95->99	CK2 PHOSPHO_SITE	PDCC000006
PS00006	136->140	CK2 PHOSPHO_SITE	PDCC000006
PS00006	183->187	CK2 PHOSPHO_SITE	PDCC000006
PS00006	189->193	CK2 PHOSPHO_SITE	PDCC000006
PS00006	256->260	CK2 PHOSPHO_SITE	PDCC000006
PS00006	445->449	CK2 PHOSPHO_SITE	PDCC000006
PS00006	463->467	CK2 PHOSPHO_SITE	PDCC000006
PS00006	546->550	CK2 PHOSPHO_SITE	PDCC000006
PS00007	364->372	TYR PHOSPHO_SITE	PDCC000007
PS00008	137->143	MYRISTYL	PDCC000008
PS00008	273->279	MYRISTYL	PDCC000008
PS00008	289->295	MYRISTYL	PDCC000008

(No Pfam data available for DKFZphmcf1\_1g13.1)

DKFZphtes3\_14g5  
-----

group: testes derived

DKFZphtes3\_14g5 encodes a novel 379 amino acid protein with strong similarity to murine cell growth regulating nucleolar protein LYAR.

The novel protein is very similar to murine Ly-1 antibody reactive clone protein (LYAR). It contains a ATP/GTP-binding site motif A (P-loop, interacts with one of the phosphate groups of a ATP/GTP nucleotide), but not the zinc finger motif and and nuclear localization signals of lyar.

No informative BLAST results; No predictive prosite, pfam or SCOP motive.

The new protein can find application in studying the expression profile of testis-specific genes.

strong similarity to cell growth regulating nucleolar protein LYAR, of mouse

complete cDNA, complete cds, EST hits

Sequenced by BMFZ

Locus: unknown

Insert length: 1503 bp

Poly A stretch at pos. 1467, polyadenylation signal at pos. 1440

```

1 CCCAGAGGTC CGACCTGGGA GGCTGGGGCT CAGAGAGCAA TGTTTGCTGT
51 CTTCCATTGG AGTGACTGAA TTTCTACATG ACGGCTTTT GACAAGACTT
101 AAAACCTGTC TTGGATAGAG AATATTTAGC CATTACCTA AAAATGGTAT
151 TTTTACATG CAATGCATGT GGTGAATCAG TGAAGAAAT ACAAGTGGAA
201 AAGCATGTGT CTGTTTGCGA AAACGTGTGA TGCTTTTCT GCATTGACTG
251 CGGTAAGAT TTCTGGGGCG ATGACTATAA AAACCACGTG AAATGCATAA
301 GTGAAGATCA GAAGTATGGT GGCAAGGCT ATGAAGGTAA AACCCACAAA
351 GCGCAGATCA AACAGCAGGC GTGGATTTCAG AAAATTAGTG AATTAATAAA
401 GAGACCCAAT GTCAGCCCA AAGTGAGAGA ACTTTAGAG CAAATTAGTG
451 CTTTGTACAA CGTTCCAGG AAAAGGGCAA AATTCAGAA TTGGATGAAG
501 AACAGTTTAA AAGTTCATAA TGAATCCATT CTGGACCAGG TGTGGAATAT
551 CTTTCTGAA GCTTCCAAAC GCGAACCATG CAATAAGGAA CAGGATCAAC
601 GGCCACTCCA CCCAGTGGCA AATCCACATG CAGAAATCTC CACCAAGGTT
651 CCAGCCTCCA AAGTGAAAGA CGCCGTGGAA CAGCAAGGGG AGGTGAAGAA
701 GAATAAAAGA GAAAGAAAGG AAGAACGGCA GAAGAAAGG AAAAGAGAAA
751 AGAAGAAGT AAAATTAGAA AACCACAGG AAAACTCAAG GAATCAGAAG
801 CCTAAGAAGC GCAAAAAGGG ACAGGAGGCT GACCTTGAGG CTGGTGGGGA
851 GGAAGTCCCT GAGGCCAATG GCTCTGAGG GAAGAGGAGC AAGAAGAAGA
901 AGCAGCGCAA GGACAGGCC AGTGAGGAAG AGGCAGCGT GGGCGCAGGG
951 AAGAGCAAGC CGAGGCACCT GGAAGTTGAA ACAGATTCTA GAAGAAGAAA
1001 GATGAAGCTC CCAGAGCATC CTGAGGGCGG AGAACAGAA GACGATGAGG
1051 CTCTGCAAAA AGGTAAATTC AACTGGAAGG GAATATTAA AGCAATCTCTG
1101 AACAGGGCCC CAGACATGA AATAACCATC AAAAGCTAA GGAAGAGGT
1151 TTTAGCTCAG TACTACACAG TGACAGATGA GCATCACAGA TCCGAAGAGG
1201 AACTCCTGGT CATCTTTAAC AAGAAAATCA GCAAGAACCC TACCTTTAAG
1251 TTATTAAAGG ACAAGTCAA GCTTGTGAAA TGAACATTG TGTATTTAAA
1301 AATTGAATCC ATTCTGCTGA CTCTTCTCT TCACTGCTGT TTATAAAATG
1351 TGTAATGAAT TCTAACAAC CAAATTTTC TTTTGAAGC TGTATTTTAA
1401 AGTTAAGAAA ATATATTTT GGTATAACTT TTATGAGAAA AATAAAATAT
1451 ATTTGGTCC AAACCTCAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA
1501 AAA
```

#### BLAST Results

-----

No BLAST result

#### Medline entries

-----

93259460:

LYAR, a novel nucleolar protein with zinc finger DNA-binding motifs, is involved in cell growth regulation.

## Peptide information for frame 3

ORF from 144 bp to 1280 bp; peptide length: 379  
Category: strong similarity to known protein  
Classification: Cell division  
Prosites motifs: ATP\_GTP\_A (60-68)

```
1 MVFFTCNACG ESKKIQVEK HVSVCNCEC LSCIDCGKDF WGGDYKNHVK
51 CISEDQKYG KGYEGKTHKG DIKQAWIQK ISELIKRPNV SPKVRELLEQ
101 ISAFDNVPRK KAKFQNMWKN SLKVHNSIL DQVWNIFSEA SNSEPVNKEQ
151 DORPLHPVAN PHAEISTKVP ASKVKDAVEQ QGEVKNKRE RKEERQKRRK
201 REKKELKLEN HQENSRNQPK KRRKKQGEAD LEAGGEEVPE ANGSAQKRSK
251 KKKQRKDSAS EEARVAGAK KRRHSEVET DSKKKMKLP EHPEGGEPEP
301 DEAPAKGKFN WKGTIKAILK QAPDNEITIK KLRKKVLAQY YTVTDEHRS
351 EEELLVIFNK KISKNPTFKL LKDKVKLVK
```

## BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphtes3\_14g5, frame 3

PIR:A40683 cell growth regulating nucleolar protein LYAR - mouse, N = 1, Score = 1410, P = 2.7e-144

SWISSPROT:YQ58\_CAEEL HYPOTHETICAL 28.5 KD PROTEIN C16C10.8 IN CHROMOSOME III., N = 1, Score = 381, P = 2.9e-35

TREMBL:AC003058\_18 gene: "F27F23.18"; product: "putative RNA-binding protein"; Arabidopsis thaliana chromosome II BAC F27F23 genomic sequence, complete sequence., N = 3, Score = 139, P = 4e-15

PIR:S70049 nucleic acid-binding protein YCR087c-a - yeast (Saccharomyces cerevisiae), N = 1, Score = 164, P = 1.4e-11

>PIR:A40683 cell growth regulating nucleolar protein LYAR - mouse  
Length = 388

## HSPs:

Score = 1410 (211.6 bits), Expect = 2.7e-144, P = 2.7e-144  
Identities = 275/388 (70%), Positives = 317/388 (81%)

```
Query: 1 MVFFTCNACGESVKKIQVEKHVSVCNCECLSCIDCGKDFWGGDYKNHVKCISEDQKYG 60
MVFFTCNACGESVKKIQVEK VS CRNCECLSCIDCGKDFWGGDYK+HVKCISE QKYGG
Sbjct: 1 MVFFTCNACGESVKKIQVEKQVSNCRNCECLSCIDCGKDFWGGDYKSHVKKCISEGQKYGG 60

Query: 61 KGYEGKTHKGDIKQAWIQKISELIKRPNVSPKVRELLEQISAFDNVPRKKAKFQNMWKN 120
KGYE KTHKGD KQAWIQKI+ELIK+PNVSPKVRELL+QISAFDNVP KAKFQNMWKN
Sbjct: 61 KGYEAKTHKGDAKQAWIQKINELIKRPNVSPKVRELLQISAFDNVPIKAKFQNMWKN 120

Query: 121 SLKVHNSILDQVWNIFSEASNSPEVNKEQDORPLHPVANPHAEIS-TKVPASKVKDAVE 179
SLKVH++S+L+QVW+IFSEAS+SE ++Q Q P H A PHAE+ TKVP++K E
Sbjct: 121 SLKVHSDSVLEQVWDFSEASSSE---QDQQPPSH-TAKPHAEMPITKVPSAKTNGTTE 176

Query: 180 QGEVKNKRRERKEERQKRRKKREKKELKLENHQENSRNQPKRRKKQGEADLEAGGEEVP 239
+Q E KKNKRRERKEERQK RK+EKKEKLENHQEN R QKPKRRK QEA EA GE+
Sbjct: 177 EQTEAKKNKRRERKEERQKNRRKKREKKELKLENHQENLRGQPKRRKKKQEAQHEAAGEDGA 236

Query: 240 EANG-----SAGRSKKKKQRKDSASEEEA----RVGAGKRRK-RHSEVETDSKKKKM 287
+ +G +G S++ R E+ A + AGKRRK +HS E+ KKKKM
Sbjct: 237 DSGPPEKKKAQGGQASEEGADRNGGPGEDRAEGQTKTAAGKRRKPKHSGAESGYKKKKM 296

Query: 288 KLPEHPEGGEPEDEAPAKGKFNWKGTIKAILKQAPDNEITIKLRKKVLAQYTVTDEH 347
KLPE PE GE +D EAP+KGKFNWGTIKA+LKQAPDNEI++KKL+KKV+AQY+ V ++
Sbjct: 297 KLPEQPEEGEAKDHEAPSKGFNWKGTIKAVLKQAPDNEISVKKLKKKVIAQYHAVMNDT 356

Query: 348 HRSEELLVIFNKKISKNPTFKLLKDKVKLVK 379
EEELL IFN+KIS+NPTFK+LKD+VKL+K
Sbjct: 357 SHHEELLAIIFNRKISRNPFTFKVLKDRVKLLK 388
```

Pedant information for DKFZphtes3\_14g5, frame 3

## Report for DKFZphtes3\_14g5.3

[LENGTH] 379  
 [MW] 43634.03  
 [pI] 9.59  
 [HOMOL] PIR:A40683 cell growth regulating nucleolar protein LYAR - mouse 1e-122  
 [FUNCAT] 04.99 other transcription activities [S. cerevisiae, YCR087c-a] 2e-11  
 [BLOCKS] BL00603D Thymidine kinase cellular-type proteins  
 [BLOCKS] BL00530C  
 [PROSITE] ATP\_GTP\_A 1  
 [KW] All\_Alpha  
 [KW] LOW\_COMPLEXITY 18.73 %

```

SEQ  MVFFTCNACGESVKKIQVEKHVSVCNRNCECLSCIDCGKDFWGDYKNHVKCISEDQKYGG
SEG  .....
PRD  cccccccccccchhhhhhhheeeccccceccccccccccccceeecccccccc

SEQ  KGYEGKTHGDIKQAWIQIKISELIKRPNVSPKVRELLEQISAFDNVPRKKAKFQNMKN
SEG  .....
PRD  cccccccccchhhhhhhhhhhhhhhccchhhhhhhhhhhcccccchhhhhhhhhhhc

SEQ  SLKVHNESILDQVWNIFSEASNSEPVNKEQDQRPLHPVANPHAEISTKVPASKVKDAVEQ
SEG  .....
PRD  cccccchhhhhhhhhhhhhhhccchhhhhhhhhccccccccccccceccccchhhhh

SEQ  QGEVKKNKREERQKKRKREKELKLENHQENSRNQPKKKKKGQADLEAGGEEVPE
SEG  ..xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx
PRD  hhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhccccccccchhhhhhhhhhhcccccc

SEQ  ANGSAKRSKKKQKQKDSASEEERVGAGRRKRHSEVETDSKKKKMKLPEHPEGGEPEP
SEG  ..xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx
PRD  cccccccchhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhccccccccchhhhhhhhhhhcccccc

SEQ  DEAPAKGKFNWGTIKAILKQAPDNEITIKLRKKVLAQYTYTDEHHRSEELLVIFNK
SEG  xxxxx.....
PRD  cccccceehhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhccccchhhhhhhhhhhhhhhhh

SEQ  KISKNPFTKLLKDKVKLVK
SEG  .....xxxxxxxxxxxxx
PRD  cccccchhhhhhhhhhhcc
  
```

## Prosite for DKFZphtes3\_14g5.3

PS00017      60->68    ATP\_GTP\_A      PDOC00017

(No Pfam data available for DKFZphtes3\_14g5.3)

DKFZphtes3\_14h21

-----

group: nucleic acid management

DKFZphtes3\_14h21 encodes a novel 648 amino acid protein with strong similarity to mus musculus RNA helicase and several RNA-dependent ATPases from the DEAD box family.

RNA helicases comprise a large family of proteins that are involved in basic biological systems such as nuclear and mitochondrial splicing processes, RNA editing, rRNA processing, translation initiation, nuclear mRNA export, and mRNA degradation. RNA helicases are essential factors in cell development and differentiation, and some of them play a role in transcription and replication of viral single-stranded RNA genomes. The members of the largest subgroup, the DEAD and DEAH box proteins, exhibit a strong dependence of the unwinding activity on ATP hydrolysis. The novel protein contains a DEAD-box and a ATP/GTP-binding site motif A (P-loop) and is a new member of this subgroup.

The new protein can find application in modulating RNA metabolism and gene expression.

strong similarity to RNA helicases

start at Bp 33 matches Kozak consensus ACNATg

Sequenced by BMF2

Locus: unknown

Insert length: 2200 bp

Poly A stretch at pos. 2166, polyadenylation signal at pos. 2140

```
1 CAACGACGTC GGACGCGCCC CTTCTTGGAA CAATGTCCCA CCACGGAGGA
51 GCTCCCAAGG CCTCTACGTG GTCGTTGCT AGTCGGCGAA GCTCGACAGT
101 GTCCCGAGCG CCAGAGAGGA GGCCGGCGGA GGAGTTGAAT CGAACAGGTC
151 CTGAGGGGATA TAGTGTCCGC AGAGGTGGTC GCTGGAGAGG CACCTCTAGG
201 CCCCCGGAGG CCGTGGCCGC TGGTCACGAG GAACGCCCGC TGTGTTTTCG
251 TTTGAAGAGC CACTTTGTTG GCGCGGTAAT CGGTCTGTGT GGGTCAAAAA
301 TAAAGAAATAT ACAAGTACA ACAACACCA CAATCCAAAT AATACAAGAA
351 CAACCAGAAT CATTAGTCAA AATTTTGGC AGCAAGGCAA TGCACACGAA
401 AGCAAAAGCA GTGATAGACA ATTTTGTAA AAAGCTAGAA GAAATATACA
451 ATTCAGAAATG CGGAATTGAT ACTGCATTCC AACCTTCTGT TGGAAAAGAT
501 GGAAGCACAG ATAACAATGT TGTTCAGGGA GATCGGCCAT TGATAGATTG
551 GGATCAAAAT AGAGAGGAAG GTTTGAAATG GCAAAAAACA AAGTGGGCAG
601 ATTTACCACC AATTAAGAAA AACTTTTATA AAGAGTCCAC TGCCACAAAT
651 GCCATGTCAA AAGTAGAAGC AGATAGTTGG AGGAAAGAAA ATTTTAATAT
701 AACGTGGGAT GACTTGAAGG ATGGGGAGAA ACGACCTATC CCCAATCCTA
751 CCTGCACATT TGATGACGCC TTCAATGTT ATCCTGAGGT TATGGAAAAC
801 ATTAATAAGG CAGGTTTCA AAAGCCAACA CCTATTCACT CACAGGCATG
851 GCCCATTGTG TTGCAAGGAA TAGATCTTAT AGGAGTAGCC CAGACTGGAA
901 CAGGAAAGAC ATTGTGTTAT TTAATGCCCT GATTATTCA TCTGGTCTCT
951 CAACCCAGCC TTAAGGTCA AAGGAATAGA CCCGGCATGT TAGTTCTAAC
1001 TCCCACTCGG GAATTAGCAC TTCAAGTAGA AGGAGAATGT TGCAAAATAT
1051 CATATAAAGG GCTTCGGAGT GTTTGTGTAT ATGGTGGTGG AAATAGAGAT
1101 GAACAAATAG AAGAGCTTAA AAAAGGTGTA GATATCATAA TTGCAACTCC
1151 CGGAAGATTG AATGATCTGC AAATGAGTAA CTTCTGCAAT CTGAAGAATA
1201 TAACCTACTT GGTTTTAGAT GAAGCAGACA AGATGTGGA CATGGGATTT
1251 GAACCCAGCA TAATGAAGAT TTTGTTAGAT GTGCGCCAG ATAGGCAGAC
1301 AGTTATGACC AGTGCTACAT GGCCTCATTC AGTTTCATCG CCGCACAAT
1351 CTTATTGAA ARAACCAATG ATTGCTATG TTGGTACATT GGATCTAGTT
1401 GCTGTAAGTT CAGTGAAGCA AATATAATT GTAACCAACG AGGAAGAGAA
1451 ATGGAGTCAC ATGCAAACTT TTCTACAGAG TATGTCATCC ACAGACAAG
1501 TCATTGTCTT CGTTTCTCGA AAAGCTGTTG CGGATCACTT ATCAAGTGAC
1551 CTAATACTTG GAAATATATC AGTAGAGTCT CTGCATGGAG ATAGAGAACA
1601 GAGAGATCGG GAGAAAGCAT TAGAGAACTT TAAACAGGCG AAAGTGAGAA
1651 TACTAATTGC AACTGATCTA GCCTCTAGAG GACTTGATGT CCATGACGTT
1701 ACACATGTCT ATAATTTTGA CTTTCCACGG AATATTGAAG AATACGTACA
1751 CCGAATAGGG CGCACGGGAA GAGCAGGGAG GACTGCTGTT TCCATTACAA
1801 CTTTGACTAG AAATGATTGG AGGGTTCCT CTGAATTGAT TAATATTCTG
1851 GAAAGAGCAA ATCAGAGTAT TCCAGAGGAG CTTGTATCAA TGGCTGAGAG
1901 GTTTGAGGCA CATCAACGGA AAAGGGAAAT GGAAGAGAAA ATGGAAAGAC
1951 CTCAGGAGAG GCCCAAGAGG TTTTATTAAT GTCTTCTGTA CTAGTGGGGT
2001 AGAGAATTCA AGATTTTFTA GAAATATAGT AAGACAGAAG TATTGGACAT
2051 GTTGGCAGTA TGAAGAGACC GGACTGATTT GACTGATTCT TAAATAATA
2101 GTGTTTGAAA ATATAGAAATC CAGTGTTTAA TACTTTCTTT AATAAAAAATA
2151 GAAGTATTTA AACTTGAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA
```

BLAST Results

-----

## No BLAST result

## Medline entries

No Medline entry

## Peptide information for frame 3

ORF from 33 bp to 1976 bp; peptide length: 648  
 Category: strong similarity to known protein  
 Classification: Nucleic acid management  
 Prosite motifs: ATP GTP A (286-294)  
 DEAD\_ATP\_HELICASE (394-403)

```

1 MSHHGGAPKA STWVVASRRS STVSRAPERR PAEELNRTGP EGYSVGRGGR
51 WRGTSRPPEA VAAGHEELPL CFALKSHFVG AVIGRGGSKI KNIQSTNTT
101 IQIQEQEPES LVKIFGSKAM QTKAKAVION FVKLEENYN SECGIDTAFO
151 PSVGKDGSTD NNWVAGDRPL IDWDQIREEG LKWKTKWAD LPPIKKNFYK
201 ESTATSAMSK VEADSWRKEN FNITWDDLKD GEKRPINPT CTFDADFQCY
251 PEVMENIKKA GFQKPTPIQS QAWPIVLQGI DLIGVAQTGT GKTLCYLMFG
301 FIHLVLQPSL KGQRNRPCLM VLTPTRELAL QVEGECKYS YKGLRSVCVY
351 GGGNRDEQIE ELKKGVDIII ATPGRLNDLQ MSNFVNLKNI TYLVLEADK
401 MLDMGFEFQI MKILLDVRPD RQTVMTSATW PHSVHRLAQ VLEKPMIVYV
451 GTLDLVAVSS VRQNIIVTTE EEKWSHMOTF LQSMSTDKV IVFVSRKAVA
501 DHLSSDLILG NISVESLHGD REQDRKAL ENFKTGKURI LIATDLASRG
551 LDVHDVTHVY NFDFFRNIEE YVHRIGRTGR AGRTGVSITT LTRNDWRVAS
601 ELINILERAN QSIPEELVSM AERFEAHQRK REMERKMERP QGRFKKPH

```

## BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphtes3\_14h21, frame 3

TREMBL:CEY54G11A\_9 gene: "Y54G11A.3"; Caenorhabditis elegans cosmid  
 Y54G11A, N = 1, Score = 1008, P = 1.1e-101

TREMBL:SPBP8B7\_16 gene: "dbp2"; "SPBP8B7.16c"; product: "p68-like  
 protein."; S.pombe chromosome II pl p8B7., N = 1, Score = 971, P =  
 9.1e-98

PIR:S13757 RNA helicase DBP2 - yeast (Saccharomyces cerevisiae), N = 1,  
 Score = 970, P = 1.2e-97

PIR:S14048 RNA helicase dbp2 - fission yeast (Schizosaccharomyces  
 pombe), N = 1, Score = 961, P = 1e-96

PIR:A57514 RNA helicase HEL117 - rat, N = 2, Score = 888, P = 7.8e-91

>TREMBL:CEY54G11A\_9 gene: "Y54G11A.3"; Caenorhabditis elegans cosmid  
 Y54G11A  
 Length = 504

## HSPs:

Score = 1008 (151.2 bits), Expect = 1.1e-101, P = 1.1e-101  
 Identities = 211/473 (44%), Positives = 298/473 (63%)

```

Query: 174 DQIREEGLKWKQTKWADLPPIKKNFYKSTATSAMSKEADSWRKENFNITWDDLKDGEK 233
      D+++E W K P I ++ YK +S + + ++
Sbjct: 23 DRLKDNFSWMK-----PIVRDLYKIPNEQKNLSPEQLQELYTNGGVMKVYPFREEST 75

Query: 234 RPIPNPTCTFDADFQCYPEVMENIKKAGFQKPTPIQSQAQWPIVLQGLDQIGVAQTGTGKT 293
      IP P +F+ AF +M I+K GF+KP+PIQSQ WP++L G D IGV+QTG+GKT
Sbjct: 76 VKIPPPVNSFEQAFGSNASINGEIRKNGFEKPSPIQSQMWPLLSSGQDCIGVSGTSGGNT 135

Query: 294 LCYLMPGFIHLVLQPSL----KGQRNRPGLMLVLTPTRELALQVEGECKYSYKGLRSVC 348
      L +L+P +H+ Q + + Q+ P +LVL+PTRELA Q+EGE KYSY G +SVC
Sbjct: 136 LAFLLPALLHIDAQLAQYEKNDEEQKPSFVLVLSPTRELAQQIEGEVKKYSYNGYKSV 195

Query: 349 VYGGNRDEQIEELKKGVDIIIATPGRLNDLQMSNFVNLKNITLVLEADKMLDMGFEP 408

```

```

Sbjct: 196 +YGGG+R EQ+E + GV+I+IATPGRL DL ++L ++TY+VLDEAD+MLDMGFE
LYGGGSRPEQVEACRGGVEIATPGRLTDLSDNGVISLASVTYVVLDEADRLDMGFEV 255

Query: 409 QIMKILLDVRPDRQTVMTSATWPHSVHRLAQSYLEKPMIVYVGTLDLVAVSSVKQNIIVT 468
I +IL ++RPDR +TSATW V +L Y KE ++ G+LDL + SV Q
Sbjct: 256 AIRRILFEIRPDRLVALTSATWPEGVRKLTDKYTKAAMVAVNGSLDLTSCSKSVTQFFFEV 315

Query: 469 TEEKEW---SHMQTFLOQSMSSD-KVIVFVSRKAVADHLSSDLILGNISVESLHGDREQR 524
+ ++ + FL + + K+I+V K +ADHLSSD + I+ + LHG R Q
Sbjct: 316 PHDSRFLRVCEIVNFLTAAHGQNYKMIIFVKSVMADHLSSDFCMKGINSQGLHGGRSQS 375

Query: 525 DREKALENFKTKGVRIIATDLASRGDLVDVTHVYNFDFPRNIEEYVHRIGRTGRAGRT 584
DRE +L ++G+V+IL+ATDLASRG+DV D+THV N+DFP +IEEYVHR+GRTGRAGR
Sbjct: 376 DREMSLNMLRSGEVQILVATDLASRGIDVDPDITHVLNDFPMDIEEYVHRVGRTRGRGRK 435

Query: 585 GVSITTLTRNDWRVASELINILERANQSIPEELVSMAERFEAHQRKREMERKMERPOGRP 644
G +++ L ND LI ILE++ Q +P++L AE++ K + R RP R
Sbjct: 436 GEAMSLFWNDRSNFEGLIQLEKSEQVDPQLRRDAEKYRL---KCQSGRDGPRPSFRN 492

Query: 645 KK 646
K
Sbjct: 493 NK 494

```

Pedant information for DKFZphtes3\_14h21, frame 3

#### Report for DKFZphtes3\_14h21.3

```

[LENGTH] 648
[MW] 72873.51
[PI] 8.84
[HOMOL] TREMBL:CEY54G11A_9 gene: "Y54G11A.3"; Caenorhabditis elegans cosmid Y54G11A 1e-
101
[FUNCAT] 04.01.04 rna processing [S. cerevisiae, YNL112w] 2e-97
[FUNCAT] 30.10 nuclear organization [S. cerevisiae, YNL112w] 2e-97
[FUNCAT] 04.05.03 mrna processing (splicing) [S. cerevisiae, YPL119c] 4e-72
[FUNCAT] 30.03 organization of cytoplasm [S. cerevisiae, YOR204w] 2e-70
[FUNCAT] 05.04 translation (initiation, elongation and termination) [S. cerevisiae,
YOR204w] 2e-70
[FUNCAT] 06.10 assembly of protein complexes [S. cerevisiae, YBR237w] 1e-61
[FUNCAT] 1 genome replication, transcription, recombination and repair [H.
influenzae, H10892] 2e-49
[FUNCAT] j mrna translation and ribosome biogenesis [H. influenzae, H10231 RNA] 1e-48
[FUNCAT] 04.99 other transcription activities [S. cerevisiae, YDL160c] 9e-45
[FUNCAT] 04.05.01.07 chromatin modification [S. cerevisiae, YMR290c] 3e-44
[FUNCAT] 09.01 biogenesis of cell wall [S. cerevisiae, YJL033w] 2e-36
[FUNCAT] 98 classification not yet clear-cut [S. cerevisiae, YOR046c] 7e-32
[FUNCAT] 30.16 mitochondrial organization [S. cerevisiae, YDR194c] 2e-28
[FUNCAT] 99 unclassified proteins [S. cerevisiae, YGL064c] 5e-10
[FUNCAT] 11.10 cell death [S. cerevisiae, YMR190c] 2e-08
[FUNCAT] 03.19 recombination and dna repair [S. cerevisiae, YMR190c] 2e-08
[FUNCAT] r general function prediction [M. jannaschii, MJ1401] 1e-07
[BLOCKS] BL00039D DEAD-box subfamily ATP-dependent helicases proteins
[BLOCKS] BL00039C DEAD-box subfamily ATP-dependent helicases proteins
[BLOCKS] BL00039B DEAD-box subfamily ATP-dependent helicases proteins
[BLOCKS] BL00039A DEAD-box subfamily ATP-dependent helicases proteins
[PIRKW] nucleus 4e-96
[PIRKW] RNA binding 3e-87
[PIRKW] DEAD box 5e-50
[PIRKW] transmembrane protein 4e-27
[PIRKW] DNA binding 3e-67
[PIRKW] recF recombination pathway 3e-10
[PIRKW] ATP 4e-96
[PIRKW] purine nucleotide binding 5e-50
[PIRKW] P-loop 4e-96
[PIRKW] hydrolase 9e-45
[PIRKW] protein biosynthesis 5e-50
[PIRKW] ATP binding 1e-61
[SUPFAM] WW repeat homology 8e-88
[SUPFAM] DEAD/H box helicase homology 4e-96
[SUPFAM] unassigned DEAD/H box helicases 7e-87
[SUPFAM] ATP-dependent RNA helicase DBP1 4e-96
[SUPFAM] ATP-dependent RNA helicase DHX1 2e-43
[SUPFAM] recQ protein 3e-10
[SUPFAM] Bloom's syndrome helicase 5e-07
[SUPFAM] translation initiation factor eIF-4A 5e-50
[SUPFAM] recQ helicase homology 3e-10
[SUPFAM] tobacco ATP-dependent RNA helicase DB10 8e-88
[PROSITE] DEAD_ATP_HELICASE 1

```

```

[PROSITE]      ATP_GTP_A      1
[PFAM]         Helicases conserved C-terminal domain
[PFAM]         KH domain family of RNA binding proteins
[PFAM]         DEAD and DEAH box helicases
[KW]           Alpha Beta
[KW]           LOW COMPLEXITY      8.49 %

```

```
SEQ      MSHHGGAPKASTWVVASRRSSTVSRAPEERRPAEELNRTGPEGYSVGGRGWRGTSRPPEA
SEG      .....XXXXXXXXXXXXXXXXXXXXX.....
PRD      CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC
```

SEQ VAAGHEELPLCFALKSHFVGAVIGRGGSKIKNIQSTTNTTIQIQEQPESLVKIFGSKAM  
 SEG .....XXXXXXXXXXXXXXXXX.....  
 PRD cccccccccchhhhhccccceeeccccccccccccccceeeccccceeeecchh

```
SEQ      QTKAKAVIDNFVKKLEENYNSECIGDTAFQPSVGKDGSTDNNVVAGDRPLI DWDQIREEG
SEG      .....
PRD      hhhhhhhhhhhhhhhhhhhhhccccccccccccccccccccccccccccccccccccc
```

```
SEQ      LKWQKTKWADLPPIKKNFYKESTATSAMSKVEADSWRKENFNI TWDDLKDGEKRPIPNT  
SEG  
PRD      cccccccccc
```

SEQ      CTFDDAFQCYPEVMENIKKAGFQKPTPIQSQAWPIVLQIGIDLIGVAQTGTGKTLCYLMPG  
 SEG      .....  
 PRD      cccccccccchhhhhhhhhhhhhcc

```

SEQ      FIHLVLQPSLKGQRNRPGLVLTPRELALQVEGECCKYSYKGLRSVCVYGGGRDEQIE
SEG      .....
PRD      eeeeeccccccccccccceeeecchhhhhhhhhhhhhhhccceeeeeccccccchhh

```

SEQ ELKKGVDII IATPGRLNDLQMSNFVN LKNITYLV LDEADKMLDMGFEPQIMKILLDVRPD  
SEG .....  
PRD hhhhceeeeeccccchhhhhhhccccccceeehhhhhhhhhhccccchhhhhhhhhhhccc

```

SEQ      RQTVMTSATWPHSVHRLAQSYLKEPMIVYVGTLDLVAVSSVKQNIIVTTEEEKWSHMQTF
SEG
PRD      cccccccccchhhhhhhhhhhhhheeeeeccccccccccccceehhhhhchhhhhhhhh

```

SEQ LQSMSSTDKVIVFVSRKAVADHLSSDLILGNISVESLHGDREQRDREKALENFKTGKVI  
SEG .....  
PRD hhhhhccccceeeeeehhhhhhhhhhhhhhhccccceeeccccchhhhhhhhhhhhhccccce

SEQ LIATDLASRGLDVHDTVHVNFDFPRNIEEYVHRIGRTGRAGRTGVSITTLTRNDWRVAS  
SEG .....XXXXXXXXXXXXX.....  
PRD eeehhhhhhccccceeeeeccccccccceeeccccccccceeeccccchhh

```
SEQ      ELINILERANQSIPEELVSMAERFEAHQRKREMERKMERPQGRPKKFH  
SEG      .....xxxxxxxxxxxxx.....  
PRD      hhhhhhhhccccchhhhhhhhhhhhhhhhhhhhhhhccccc
```

Prosites for DKFZphtes3 14h21.3

PS00017	286->294	ATP_GTP_A	PDOC00017
PS00039	394->403	DEAD_ATP_HELICASE	PDOC00039

Pfam for DKFZphtes3\_14h21.3

HMM_NAME	DEAD and DEAH box helicases	
HMM	*GlpPwILrNIyeMGFEKPTPIQQAIPiILEGRDVMCACAQTGSGKTAAF P++++NI++++GK KPTPIQQA+PI+L+G D+++ AQTG+GKTT++	
Query	248 QCYPEVNMENIKKAGFGKPTPIQQAQPIVLQGLDILGIVAGTGTGKTLCY	296
HMM	lPMLQHIDwdPwpqpQd..PrALiAPtRELAMQIEEcRkFgkHMng L+P ++H+ +P +++ Q+ P +L+L+PtRELAtQ++ EC K+++	
Query	297 LMPGFIHLVLQPG-SLKGQNRNFGMLVLTPTTRELAtQVEGCKSYK-G-	343
HMM	IRImCIYGGtNRMDQMRNLerGpPHIViATPGRLIDHIERgtldLDRIeM +R++C+YGG N ++G++L++G+ +I+IATPGRL D+ +++ ++L++I++	
Query	344 LRSVCYGGGNLRDEIQELKKGVD-IIiATPGRLNDLQMSFNLNKITY	392
HMM	LVMDEADRLMDMGFDIQIRImrQIPmpwNRQTMMSFATMPdeIQELARR LV+DEAD+LMDMGF+QI+I++ ++ +QTM+M SAT+P ++ +LA	
Query	393 LVLDEADRLMDMGFEPQIMKILLDVR--pDRQTM+SATWPSVHRLAQs	440



```

HMM          FMRNPIRInId.MdElTtnEnIkQwYiyVerEMWKfdCLcrLIe*
          +++++P  + ++  D  +++  +KQ +I+  E++K +  +++++
Query        441 YLKEPMIVYVGTLDLVAVS-SVKQNIIVTT-BEEKWSHMQTFLQ  482

HMM_NAME     KH domain family of RNA binding proteins

HMM          *rIiIPedhMGMIIGKGGsNIRqIREEYgvrINIPdecCeDstdRIITit
          + + +++++G++IG+GGS I++I++ +++++I I++E+ +  + + I
Query        71  CFALKSHFVGAVIGRGGSIKNIQSTTNTTIQIIQE-Q-P---ESLVKIF  115

HMM          G*
          G
Query        116 G  116

HMM_NAME     Helicases conserved C-terminal domain

HMM          *EileeWLknl....GlrvmYIHGdMpQeERdeIMddFNnGEynVLicTD
          + +++ L+ +  +I+V ++HGD++Q+R+++++F++G+ ++LI+TD
Query        497 KAVADHLSSDLILGNISVESLHGDREQRDREKALENFKTKVRILIATD  545

HMM          VggRGIDIPdVNVHVINYDMPWNPEqYIQRIGRTgRIG*
          +++RG+D+ DV HV+N+D+P+N+E Y++RIGRTGR+G
Query        546 LASRGLDVHDVTHVYNFDFPRNIEEYVHRIGRTGRAG  582

```

DKF2phtes3\_14p14

group: testes derived

DKF2phtes3\_14p14 encodes a novel 159 amino acid protein without similarity to known proteins.

No informative BLAST results; No predictive prosite, pfam or SCOP motif.

The new protein can find application in studying the expression profile of testis-specific genes.

unknown

complete cDNA, complete cds, few EST hits

Sequenced by BMF2

Locus: unknown

Insert length: 3969 bp

Poly A stretch at pos. 3948, polyadenylation signal at pos. 3927

```
1 GAAGCCAGG CTCTCCTTAG TTGACTGTGT GTTAATCACC CAGCAATTC
51 ATTACTCAAC AGCTCTCCAG AGTTGCACAT TACAGCTGGG GTAGAAATGG
101 GGTGCTGAAG GCCAGGCAGA GCATTGGCT GTAGGGAGGC CGATCCTCCT
151 CGGGCCTGTT ACCGGCCGGT CTTTGTCTT AGACCTGGGG TTCTGGGCT
201 CACGGATTCC AAGGAATGGA ACCTTGGGCC ATGCTGTGA ACAGCTCTA
251 TGTGATGAC CCAGACAAGG ACAGCGGTGG CAAGATCGAC GTGACTCTGA
301 ACATCAGTTT ACCCAATCTG CACTGCGAGT TGGTTGGGCT TGACATTGAG
351 GATGAGATGG GCAGGCACGA AGTGGGCCAC ATCGACAAC CCATGAAGAT
401 CCCGCTGAAC AATGGGGCAG GCTGGCCCTT CGAGGGGCAG TTCAGCATCA
451 ACAAGGTATG GAAGCCCTGC CTCAGCCCTT TCTACCTGCT CCCCTTCTCT
501 GCTGTCTCCC CGCTCCCTGG AAACCTGGTTG TGGAGGCAC CACTCGACCT
551 GACCCCTGACA CAGCCCCCAG CAAGCGAGGG TTCTGTGCCA GCTGCCTGGC
601 CGTTCCTGCT GAGAATCTGG ATGGGGGTCC AGGCTCCCTG GGGTTTTAAG
651 CCCCTGATGG CTGGTTCAGG AAGGAGCTAC TCTTCTCTCC AGTGAGGGGG
701 ACAATGATGA GAAGACCTGA GGATTTCAG CCCCAGCCC TGGGTTCAAG
751 TCCAGCTCT ACCCTTCTTT GGGCCCTACA AGTCACCTGA CCCATCTTAG
801 GCTGAGGGTG TGATGGCGAT AATAGTATCA CGATACCACT CACTTCACAA
851 AGTTTGTGTG GGGATTAAAT GAGCTAATGC AGATTCAATC ATTGAGAAAA
901 ATTTTGAAT GGCACGTTCT GTGTTCCAGG GTCGGTGATA GGCTCTGGGG
951 CAGCGTTCCT GGGCTGGTGG GGCTCCCAT TCGGTAGAGG GAGACAGTCT
1001 ACAAAACCAGA AAGCATCAGG GATGCTAAGT GCAGTGATGA GGAATAAAGC
1051 CAAGGGGAGT GAGATGAGGT GGGCTTGAAA GTACCTTGTC CGCTCAGAAG
1101 GACCAATCAA GGTTCACGTG TGTTTGTCC TCAGAACCA GAGCTTCAGA
1151 TCCTAAGTCA AGTGGGTGAA CGCAGTGCCC TTGGGAGGGC CGAGGCACCC
1201 GGTGGCAGCT GGCAGGGTTT TGCTCAGCAC GTGCCGGCTC TCCTCGAAGC
1251 TCGGTACTGT CACAGTGGAG CCTCTCAACA ACCTGTGAG CGACACCAAT
1301 TTGACAGGTT AGGATGCTGG GGGCCAGAGA GGTAAAGTGT CTTGCCCGAG
1351 GTCACACAGC TATCTGCATG TCCCACAAC CTCTTCCCA GCCCCAGCCA
1401 AACTGAGCCA CTGGCCACTC CTGGCTTCTC CTTGTCCCTC CTGCAGCCTC
1451 TGCTCAGAAC GCCCTTCTCT CAGACCTGA CACCTGAGCT GGGGTTGCAA
1501 AGTCACTGGC CACATCCAGC CCAAAGATAA ATTTGTTTG TCCAGTATAG
1551 CATTTAACTG CATCAGAACC AGTATGAAA GACCAGGAAT CCAGATTTC
1601 GGCTTTTAAA AGTCAGAGGC TCTCACTACA CTGGGTCGT GTTCCCGCTA
1651 TGACAATGAC CTGGCACCAG TGGGCAGTGT TCCCTTTAG AGAGGGTGTG
1701 TGCTGTCCCT TCCCACAGTC CCTGGCAGGC GGCTGGAAGG CCAGGCCCTG
1751 TCATCTGTCA AGCAGGGTGG ACTTCTTACG TGACAGTTCA GGGCTCCCTT
1801 AAGTGCTAAA GCAGAAGCTG CAAGGCTTTC TTAAGGTTTC GAGTGTGCT
1851 GGGAGAAATC TGCTGCATGT TGTGGGTTAA AGGGAGTCTC TCACAGCCCC
1901 AGGCCCTCAG GAGGAGGAGA TACCAGGAGG CAGGGATGCT GGGGGTCTGT
1951 GTTCACTGGG GGCTCTCTCT GCCATGAGC TGCCACACAG CACCTTTGCC
2001 ATGCCCCGTA ATTTGGATT TATGGTGGTT GTGATGAAA GCCATTTGAG
2051 GGTTTTGAAC AGGGAGGCAA TGTAAATCAGA TTTATGCCTT AGAAGTGGAC
2101 TATCCAATAG GTTGCCAGCA GCCACATAAG GCTATTTAAA TTAATTCAAA
2151 TTAATGTAC AATTCACTCA CTCATTCTCA TCAACCACAT TTCAAGTGCT
2201 CAAAGCCAGC TGCTGGCTAG GGGCCACAGC GTTAGACAGT GCAGAGAGAA
2251 AGCACTTCCA TCGCTGAGGA AAGTTCTGCT GGACCGACA CCCTTAGAAG
2301 GATGGCTCTG GTGGCCGGGC GCGGTGGCTC AAACCTGTAA TCCCAGCACT
2351 TTGGGAGGCC GAGGTGGGTG GATCAGGAG TCAGGAGATC GAGACCATCC
2401 CGGCTAACAT GGTGAAACCC TGCTCTACT AAAAATACAA AAAAAACAA
2451 AATTAGCCGG GCGTGGTTGC GGGCACCTGT AGTCCAGCT ACTCAGGAGG
2501 CTGAGGCGGG AGAATGGCAT GAACCCGGGA GGTGGAGCTT GCAGTGAGCC
2551 AAGATCGTAC CACTGCACCT CAGTCTGGGC GACAGAGTGA GACTCCATCT
2601 CAAACAAAC AAAAAGGA TGGGGCTGG CTGGAGAGGG TGGCAGGCAG
2651 TGGTTGTGGC AGTGGAGCTG GGGAGATGTG GTCGGATTAG GGAGGTAGAA
2701 TCAATAAGAC TCAGTGAAGA ATCGGATGTG GGGGTAAGGG CACATGTGGA
```

```

2751 AGCAAGAAA CCTTTGACGT CTTTGTCTTG ACAACCGGGT GGTCTGTGT
2801 CTAGACATGG AAGCTTAGAA AAGCCTGGAG TCTGTGGGAA GTAGGTAGGG
2851 CTGGGCACTG GTCATTCCAC TCTGGTTTCC TTTGGGGTTC CCATTAGGTG
2901 TCTACAGGGA GAGGTGAAAT TGGAAATTGG AGGTGTGGAG AGTTCAGGAG
2951 AGGGTTCTGG ACCACAGATG TTGAGTGGG AGTCATTAGT GAATAGATGA
3001 TGTGGAAAGT CATGGGTCTT CAGAGTGGGG GTCCTTAAG CCTCCAGGCC
3051 AGCAGCATCA GCATCACCTG GGAGATTGTT AGGAATGCAG ATTCTCAGGC
3101 CCCCCTAAGA CCCACCGACT CTGTGCTAGA ACAAGCGCCC CTCAGAGATT
3151 CTGATGCCAC TGAAGTTTGA GGAGCATTGG TTTAAGCAAG ATTACCTACG
3201 GAGAGGCTGT AGATCCGTGT TCTAAACCTG GGGTCCACAG ACACCCCCAA
3251 GAAGAGCGGA TTGAATGCAA GAGATCTATG AAGTTGGATG GGGGAAAAAT
3301 TGACATCTTT ATTTTGTGCTA AACTCGATCT AAAGTTTAGC ATTTCCATCT
3351 GCGATGAATG TAGGCCACAA ACCACAGTAG TATTAGCAGT GCCTGGGACC
3401 TCCTCAACAA CAGAAATTGC CGGTATTAT AGCACGTTAC AGTTGTTGCA
3451 GATAATTTC AGAGACTGTT TATATGCACC ACTGTTTAA AATTACGGTG
3501 ATTGGCCAGG TGCAGTGGCT CACACCTGTA ATCCCAGCAC TTTGGGAGGC
3551 CAAAGTGGGT GGATCACTTG AGGAGTTCAA GACCAGCCTG GTCAACATGT
3601 CAAAACCTGT TATCTACAAA AAAATACAAA AGTTAACCAA GCCTATGCTT
3651 GTAGTCACAG CTACTCGGGA GGCCGAGGTG GGAGGGCTT CTGAGCCAGC
3701 GGAGGTAGAG GCTTCAGTGA GCTGAGATCG CACCACCACA CTCCAGCCTG
3751 GGTGACAGAG TGAACCCCTT AATCAATCAG TCAATAAAAA TTACAGTAAT
3801 TATTAGACCC ACCACTAGGT CATCTTATT GATGCATCAG TAAAGCAGCA
3851 TATTCAAATG TGGATTTTAA AATATTTTAA TTACTATTTA AATATCTCTT
3901 TACTTTGTAA TCCTATGCAT TTTACGCATT AAAACATTTT AAGCATTTAA
3951 AAAAAAAAAA AAAAAAAAAA

```

# BLAST Results

No BLAST result

## Medline entries

No Medline entry

## Peptide information for frame 3

ORF from 216 bp to 692 bp; peptide length: 159  
 Category: putative protein  
 Classification: no clue

```

1 MERWAMRVNE LYVDDPKDSD GGGKIDVSLNI SLPLHCELV GLDIQDEMGR
51 HEVGHIDNSM KIPLNNGAGC RFEGQFSINK VWKPCLSPFY LLPFAVSP
101 PGNWLWRHSL DLTLTQPPAS EGSCPAAWPF LLRIWMGVQA PWGFKPLMAG
151 SGRSYSSLQ

```

## BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphtes3\_14p14, frame 3

No Alert BLASTP hits found

Pedant information for DKFZphtes3\_14p14, frame 3

## Report for DKFZphtes3\_14p14.3

```

(LENGTH) 159
(MW) 17778.55
(pI) 5.74
[FUNCAT] 99 unclassified proteins [S. cerevisiae, YAL042w] 5e-04
(KW) Alpha_Beta

```

```

SEQ MERWAMRVNELYVDDPKDSDGGKIDVSLNISLPLHCELVGLDIQDEMGRHEVGHIDNSM
PRD ccchhhhhhhcccccccccccccccccccccccccccccccccccccccccccccccc
SEQ KIPLNNGAGCRFEGQFSINKVWKPCLSPFYLLPFAVSPPLGNWLWRHSLDLTLTQPPAS
PRD eeeeeeeeecccccccccccccccccccccccccccccccccccccccccccccccccccc

```

WO 01/12659

PCT/IB00/01496

SEQ EGSCPAAWPFLLRIWMCVQAPWGFKPLMAGSGRSYSSLQ  
PRD cccccchhhhhhhhhcccccccccccccccccccc

(No Prosite data available for DKFZphtes3\_14p14.3)

(No Pfam data available for DKFZphtes3\_14p14.3)

DKFZphtes3\_l4p7

-----

group: testes derived

DKFZphtes3\_l4p7 encodes a novel 702 amino acid protein with very weak similarity to kinesin associated protein KAP3.

No informative BLAST results; No predictive prosite, pfam or SCOP motife.

The new protein can find application in studying the expression profile of testis-specific genes.

weak similarity to kinesin associated protein KAP3

complete cDNA, complete cds, few EST hits

Sequenced by BMFZ

Locus: unknown

Insert length: 2497 bp

Poly A stretch at pos. 2424, polyadenylation signal at pos. 2400

```
1  GGAATCCAAA  GAAACAGTTA  TGATGGGGGA  CTCTATGGTG  AAAATAAATG
51  GGATTTATTT  AACAAAATCA  AATGCTATTT  GCCACTTAAA  GAGTCACCCA
101  CTTCAGCTAA  CTGATGATGG  AGGCTTCAGT  GAAATAAAGG  AGCAAGAAAT
151  GTTCAAAGGA  ACAACATCTT  TACCATCTCA  TCTCAAGAAT  GGAGGGGACC
201  AGGGGAAGAG  ACATGCGAGG  GCCTCATCAT  GCCCCAGTAG  CTCAGACCTG
251  AGCAGGCTGC  AAACCAAAGC  AGTCCCAAAA  GCTGACCTGC  AAGAAGAGGA
301  CGCAGAAATA  GAAGTAGACG  AAGTCTTTTG  GAATACAAAG  ATTGTACCGA
351  TTTTGCCTGA  ATTAGAAAAG  GAAGAAAACA  TTGAAACGGT  TTGTGCTGCT
401  TGCACACAA  TTATCATATG  TTTAGAGGAA  GGAACATGTC  TTGGAATAAA
451  ATTTAAGGGA  AGAAGTATTC  TCCTGAAGAC  CCTGTGTAAA  CTAGTTGATG
501  TTGGTTCAGA  CTCGCTCAGC  CTAAACTTGG  CAAAAATAAT  TCTAGCACTT
551  AAGTGAGTGA  GAAAGAATCT  TCTTAATGTC  TGCAAACTTA  TATTTAAAT
601  TAGCAGGAAT  GAGAAAGATG  ATCTTTGAT  TCAAAATGAC  AGCATTCTGG
651  AATCATTATT  GGAGGTACTA  AGAAGTGAA  ACCTGCAAA  TAACATGGAA
701  GCTTTTTTAT  ACTGTATGGG  GCTATATAAG  TTCATTCTG  GAAATCTGGG
751  ATTTCTTAAT  GAAATGATCA  GCAAAGGTGC  TGTGGAATA  CTGATAAATT
801  TGATAAAACA  AATAAATGAG  AACATCAAGA  AATGGGTAC  ATTTTGGCT
851  AATTCGGGCC  ACTTGCTAGT  CCAGGTGACT  GCTACATTGA  GAAACTTGGT
901  TGATTCATCA  TTAGTAAGAA  GTAAGTCCCT  AAACATCAGT  GCCCTTCCCC
951  AGCTCTGCAC  GGCAATGGAA  CAGTACAAGG  GTGACAAGGA  CGTCTGTACC
1001  AATATTGCCA  GAATATTCAG  CAAACTTACT  TCTTACCGTG  ACTGCTGCAC
1051  AGCCTTGGCC  AGCTATTCCA  GATGTTATGC  CTTATTCTG  AATCTAATTA
1101  ACAATACCA  GAAGAAGCAG  GATTTAGTCG  TCCGTGTTGT  TTTTATCTTT
1151  GGCAACCTGA  CGGCAAAAA  TAACCAGGCT  CGTGAACAA  TTTCCAAAGA
1201  GAAAGGAGC  ATCCAAACTC  TGCTGTCAAT  ATTCCAGACG  TTCCATCAGC
1251  TGGATCTGCA  TTCCAGAAG  CCGTGGGCC  AACGAGGCGA  GCAGCACAGG
1301  GCGCAGAGGC  CGCCGTGAGA  GGCAGAGGAC  GTGCTCATCA  AGCTGACTCG
1351  TGTGCTGGCC  AACATTGCCA  TCCACCCGGG  CGTGGGCCCG  GTGCTGGCCG
1401  CCAACCCGGG  GATAGTGGC  CTGCTCCTGA  CCACGCTGGA  ATACAAGTCA
1451  CTTGATGATT  GTGAGGAGCT  GGTGATCAAT  GCTACAGCGA  CAATCAACAA
1501  TTTATCTTAC  TACCAAGTGA  AGAATTCAT  AATTCAGAC  AAAAGCTAT
1551  ATATTGCTGA  ATTGCTCTTA  AAGCTTCTTG  TCAGTAACAA  CATGGATGGA
1601  ATCCTGGAGG  CTGTGCGTGT  TTTGCGAAAT  CTCTCCACAG  ACCATGATGT
1651  CTGCGATTTC  ATTGTGCAGA  ACAATGTCCA  CAGGTTTCATG  ATGGCGCTGC
1701  TGGATGCTCA  GCATCAGGAT  ATCTGCTTTT  CTGCTGTGG  TGTCTCTCTC
1751  AATCTCACTG  TGGATAAAGA  CAAGCSTGTC  ATCTTGAAAG  AAGGAGGTGG
1801  CATTAAAAAG  TTAGTGGAGT  GTTAAAGAGA  TTTGGGTCTT  ACTGATTGGC
1851  AGCTGGCCTG  CTTGGTTTGT  AAAACTTTAT  GGAACCTCAG  TGAAAACATC
1901  ACTAATGCTT  CGTCATGTTT  TGGAAATGAA  GACACCAACA  CACTCTTACT
1951  CTTGCTCTCA  TCATTTTAT  ATGAAGAACT  AGCACTGGAT  GGCAGTTTGT
2001  ATCCAGACCT  AAAAACTAT  CACAAACTCC  ATTGGGAAAC  AGAATTCAAA
2051  CCTGTGGCAC  AGCAGCTTCT  AAACCGAATT  CAGAGACATC  ACACCTTCTC
2101  GGAACCCCTG  CCCATCTCCC  CTTCTAACA  TGATGCAGAT  TAACAGTAGA
2151  AACGAGAACT  CACGTCTCCC  TCATTCTTAA  GAACGTGTAA  CAACGTGAA
2201  CATTTTTTTC  AGCATTAACA  AATGTGGAAA  GTTTTCAAG  AACTGGTTTT
2251  AGTGAGTAGC  TGAAGTATTT  TTTAAATTA  AGCATTCTCT  CTTGTTAGGT
2301  ATTATGAAAT  AATGAATATA  CACATTATAT  TTCCTGTGA  GAGAAATGTA
2351  AGATCAAAAT  ATGTGCATTT  TCAAGTAAAT  GACTTTTCT  TCTATCTCT
2401  ATTAACAAT  TTAGTCTAG  TCTTAAAAA  AAAAAAATA  AAAAAAATA
2451  AAAAAAATA  AAAAAAATA  AAAAAAATA  AAAAAAATA  AAAAAAATA
```

BLAST Results

-----

No BLAST result

Medline entries

No Medline entry

Peptide information for frame 2

ORF from 20 bp to 2125 bp; peptide length: 702  
Category: putative protein

```
1 MMGDSMVKIN GIYLTGSNAI CHLKSHPLQL TDDGGFSEIK EQEMFKGTTS
51 LPSHLKNGGD QGKRHRARASS CPSSDLSRL QTKAVPKADL QEEDAIEVD
101 EVFWNTRIVP ILRELEKEEN IETVCAACTQ LHHALEEGNM LGNKFKGRSI
151 LLKTLCKLVD VGSDSLSLKL AKIILALKVS RKNLLNVCKL IFKISRNEKN
201 DSIQNDISIL ESLLLEVLRSB DLQTNMEAFI YCMGSIKFIS GNLGFLNEMI
251 SKGAVEILIN LIQINENIK KCGTFLPNSG HLLVQVTATL RNLVDSLSVR
301 SKFLNISALP QLCTAMEQYK GDKDVCTNIA RIFSKILTSYR DCCTALASYS
351 RCYALFLNLI NKYQKQDLV VRVVFILGNL TAKNNQAREQ FSKEKGSIQI
401 LLSLFTQTHQ LQLHSQKPVG QRGEQHRARQ PPSEAEDVLI KLTRVLANTA
451 IHPGVGPVLA ANPGIVGLLL TTLEYKSLDD CEELVINATA TINNLSSYYQV
501 KNSIIQDKKL YIAELLKLL VSNMMDGILE AVRUVGNLSQ DHDVCDFIQV
551 NNVHRFMAL LDAQHODICF SACGVLLNLT VDKDKRVILK EGGGIKKLVD
601 CLRD LGPTDW QLA CLVCKTL WNFSENITNA SSCFGNEDTN TLLLLSSFL
651 DEELALDGSF DPDLKNYHKL HWETEFPKVA QQLLNRIQRH HTFLEPLPIP
701 SF
```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphtes3\_14p7, frame 2

TREMBL:MMD367\_1 product: "KAP3B"; Mus musculus mRNA for KAP3B,  
complete cds., N = 2, Score = 97, P = 0.00039

>TREMBL:MMD367\_1 product: "KAP3B"; Mus musculus mRNA for KAP3B, complete  
cds.

Length = 772

HSPs:

Score = 97 (14.6 bits), Expect = 3.9e-04, Sum P(2) = 3.9e-04  
Identities = 45/163 (27%), Positives = 77/163 (47%)

Query: 442 LTRVLANIAIHPGVGPVLAANPGIVGLLLTTLEYKSLDDCEELVINATATINNLSSYYQVK 501  
L +++ NI+ H G P VG L + S D+ EE VI T+ NL+ +

Sbjct: 483 LMKMIRNISQHDG--PTKNLFIDYVGDIAAQI---SSDEEEFVIECLGTLANLTIPDL 537

Query: 502 -NSIIQDKKLYIAELLKLLVSNMMDG-ILEAVRVFGNLSQDHDVCDFIQNNVHRFMMA 559  
++++ KL + L KL D +LE V + G +S D + + + + ++

Sbjct: 538 WELVLKEYKL-VFLLKDKLPGAAEDDLVEVIMIGTVSMDDSCAALLAKSGIIPALIE 596

Query: 560 LLDAQHQDICFSACGVLL---NLTVDKDKR-VILKEGGGIKKLVDCIAD 604  
LL+AQ +D F C ++ + + R VI+KE L+D + D

Sbjct: 597 LLNAQQEDDEF-VCQIIYVFYQMVFHQATRDVIKETQAPAYLIDLMHD 644

Score = 77 (11.6 bits), Expect = 3.9e-04, Sum P(2) = 3.9e-04  
Identities = 42/178 (23%), Positives = 82/178 (46%)

Query: 169 KLAKIILALKVSRKNLLNVCK-LIFKISRNEKNDSLIQNDISLESLLLEVLRSIDLQTNME 227  
K K L V ++ LL V L+ + + + + ++N +I+ L++ L + N E

Sbjct: 263 RTFKKYQGLVVKQEQLLRVALYLLNLAEDTRTELKMRKNKNIHMLVKALDRD---NFE 318

Query: 228 AFLYCMGSIKFISGNLGFNEMISKGAVEILINLIQINENIKKCGTFLPNSGHLVQVT 287  
+ + +K +S + N+M+ VE L+ +I +E++ L + +

Sbjct: 319 LLILVVSFLKKLSIFMENKNDMVEMDIVEKLVKMIPEHEDEL-----LNITLR 366

Query: 288 ATLRLNVDSLSVRSKFLNISALPQLCTAM--EQYKGDQDVCT--NIARI--FSKILTSYRD 341  
L D+ L R+K + + LP+L + E YK +C +I+ F + +Y D

Sbjct: 367 LLLNLNLSFDGL-RNKMVQVGLLPKLTALLGNENYK-QIAMCVLYHISMDDRFKSMFAYTD 424

Query: 342 CCTAL 346  
C L  
Sbjct: 425 CIPQL 429

Score = 69 (10.4 bits), Expect = 2.6e+00, Sum P(2) = 9.2e-01  
Identities = 35/146 (23%), Positives = 70/146 (47%)

Query: 512 IAELLKLLVSNMMDGILEAVRVFGNLSQDHDVCD FIVQNNVHRFMMALLDAQHQDICFS 571  
I +L+K L +N + ++ V LS + + +V+ ++ ++ ++ +H+D+  
Sbjct: 304 IVHMLVKALDRDNFELLILVVSFLKKLSIFMENKNDMVEKDIVEKLVKMIPCEHEDLLNI 363

Query: 572 ACGVLLNLTVDKDKRVILKEGGGIKKLVDCRLDGLPTDW-QLACLVCCKTLWNFSENITNA 630  
+LLNL+ D R + + G + KL L G ++ Q+A +C L++ S +  
Sbjct: 364 TLRLLLNLSFDTGLRNKMVQVGLLPKLTALL---GNENYKQIA--MC-VLYHISMD-DRF 416

Query: 631 SSCFGNEDT-NTLLLLSSFLDEELALD 657  
S F D L+ +L DE + L+  
Sbjct: 417 KSMFAYTDCIPQLMKMLFECSDERIDLE 444

Score = 68 (10.2 bits), Expect = 3.2e-03, Sum P(2) = 3.2e-03  
Identities = 18/58 (31%), Positives = 30/58 (51%)

Query: 190 LIFKISRNEKN-DSLIQND SILESLEVLRS-----DLQTNMEAFLYCMGSIKIFISG 241  
LI +++RN N + L+ N++ L +L VLR + +L TN+ +C S G  
Sbjct: 155 LILQLARNPDNLEELLNETALGALARVLRDQKQSVELATNIYIFFCFSSFSHFHG 212

Score = 65 (9.8 bits), Expect = 6.4e+00, Sum P(2) = 1.0e+00  
Identities = 26/122 (21%), Positives = 53/122 (43%)

Query: 283 LVQVTATLRNL----VDSSLVRSKFLNISALPQLCTAMEQYKGDKDVCTNIARIFSKLTS 338  
+++ TL NL +D LV ++ +P L ++ +D+ + I S  
Sbjct: 521 VIECLGTLANLTIPDLWELVKEY---KLVPELKDCLKPFGAEDDLVLEV-VIMIGTVS 576

Query: 339 YRDCTALASYSRCYALFLNLINKYQKKQDLVVRVVFILGNLTAKNNQAREQFSKEKGS 398  
D C AL + S + L+N Q+ + V +++++ + + R+ KE +  
Sbjct: 577 MDDSCAALLAKSGIIPALIELLNAQEDDEFVCQIIVFYQMVF-HQATRODVIKETQAP 635

Query: 399 QTLTSL 404  
L+ L  
Sbjct: 636 ATLIDL 641

Score = 65 (9.8 bits), Expect = 6.4e+00, Sum P(2) = 1.0e+00  
Identities = 44/177 (24%), Positives = 79/177 (44%)

Query: 481 CE-ELVINATATIN-NLSYYQ-VKNSIIQDKKLYIAELLKLLVSNMMDGILEAVRVFGN 537  
CE E ++N T + NLS+ ++N ++Q + + L LL + N I A+ V +  
Sbjct: 355 CEHEDLNITLRLLLNLSFDTGLRNKMVQ---VGLLPKLTALLGNENYKQI--AMCVLYH 409

Query: 538 LSQDHDVCD-FIVQNNVHRFMMALLDAQHQDICFSACGVLLNLTVDKDKRVILKEGGGIK 596  
+S D F + + + M L + + I +NL +K ++ EG G+K  
Sbjct: 410 ISMDDRFKSMFAYTDCIPQLMKMLFECSDERIDLELISFCINLAANKRNVLICEGNGLK 469

Query: 597 KLVDCLRLDGLPTDWQLACLVCCKTLWNFSENITNASSCFGNEDTNTLLLLSSFLDEELAL 656  
L+ R L D L+ K + N S++ + F + L +SS +EE +  
Sbjct: 470 MLMK--RALKLKD---PLLMKMRNISQHDGPTKNLF-IDYVGDLAAQISSDDEEEFVI 522

Query: 657 D 657  
+  
Sbjct: 523 E 523

Score = 61 (9.2 bits), Expect = 1.6e-02, Sum P(2) = 1.6e-02  
Identities = 20/66 (30%), Positives = 34/66 (51%)

Query: 304 LNISALPQLCTAM-EQYKGDKDVCTNIARIFSKLTSYRDCTALASYSRCYALFLNLINK 362  
LN +AL L + E +K ++ TNI IF +S+ + Y + AL +N+I+  
Sbjct: 171 LNETALGALARVLRDQKQSVELATNIYIFFCFSSFSHFHGLITHY-KIGALCMNIIDH 229

Query: 363 YQKKQDL 369  
K+ +L  
Sbjct: 230 ELKRHEL 236

Pedant information for DKF2phtes3\_14p7, frame 2

Report for DKF2phtes3\_14p7.2

[LENGTH] 708  
[MW] 79266.35  
[pI] 6.57

[FUNCAT]	30.25 vacuolar and lysosomal organization	[S. cerevisiae, YEL013w]	3e-04
[FUNCAT]	06.04 protein targeting, sorting and translocation	[S. cerevisiae, YEL013w]	
[3e-04]			
[FUNCAT]	09.25 vacuolar and lysosomal biogenesis	[S. cerevisiae, YEL013w]	3e-04
[BLOCKS]	BL00923F Aspartate and glutamate racemases proteins		
[BLOCKS]	BL00288B Tissue inhibitors of metalloproteinases proteins		
[PROSITE]	MYRISTYL	9	
[PROSITE]	AMIDATION	1	
[PROSITE]	CK2_PHOSPHO_SITE	12	
[PROSITE]	PKC_PHOSPHO_SITE	7	
[PROSITE]	ASN_GLYCOSYLATION	11	
[KW]	Alpha Beta		
[KW]	LOW COMPLEXITY	7.49	%

```
SEQ      ESKETVMMGDSMVKINGIYLTKSNAICHLKSHPLQLTDDGGFSEIKEQEMFRGTTSLSH
SEG      .....
PRD      ccccccccccceccccccccccccccccccccccccchhhhhhcccccccc
```

```

SEQ      LKNGGDQGRHARASSCPSSSDLRLQTKAVPKADLQEEDAEIEVDEVFNTRIVPILRE
SEQ      .....XXXXXXXXXXXX
PRD      CCCCCCCHHHHHHCCCCCCHHHHHHCCCCCHHHHHHHHHHHCCCCCCEHHHHHH

```

SEQ LEKEENIETVCAACTQLHHALEEGNMI.GNKFGRSILLKTLCKLVDVGSDSLSLKLAKI I  
 SEG ..... xxxxxxxxxxxx  
 PRD hhhhhcchhhhhhhhhhhhhhhhhhhccccccccccccchhhhhheeeccccchhhhhhhh

```

SEQ      LALKVSRKNLLNVCKLIFKISRNEKNDSLIQNDSILESLEVLRSDDLQTNMEAFLYCMG
SEG      xxxx.....
PRD      hhhhhhhhhhhhhhhhhhhhhccccccccccccccccccccchhhhhhhhhhhccchhhhhhhhhhhcc

```

SEQ SIKFISGNLGLNEMISKGAVEILINLIKQINENIKKCGTFLPNSGHLLVQVTATLRNLV  
 SEQ .....  
 PRD cccccccchhhhhhchhhhhhhhhhhhhcCCCCCCCCCCCCCeeeeehhhhhhhh

SEQ DSSLVRSKFLNISALPQLCTAMEQYKGDQVCTNIARIFSKLTSYRDCCTALASYSRCYA  
 SEQ .....  
 PRD cccchhhhhheeeccchhhhhhhhhhhcccccceehhhhhhhhhhhccchhhhhhhhhhh

SEQ LFLNLINKYQKKQDLVVRVVFILGNLTAKNNQAREQFSKEKGSIQTLLSLFQTFHQDLH  
SEG .....  
PRD hhhhhhhhhhhhhhhhheeeeeeeccccccchhhhhhhhhhhhhc

SEQ SQKPVGQRGEQHQRAQRPPEAEDVLIKLTRVLANIAIHGVPVLAANPGIVGLLLTLE  
 SEQ .....  
 PRD cccccccccccccccccccchhhhhhhhhhhhhhhcccceccccchhhhhhhh

SEQ YKSLDDCEELVINATATINNLSYYQVKSIIQDKKLYIAELLKLLVSNMMDGILEAVRV  
 SEG .....XXXXXXXXXXXXXXXXX.....  
 PRD hccccchhhhhhhhhheeeccccccceeehhhhhhhhhhhhhhhhhhcccccchhhhhhhh

```
SEQ      FGNLSQDHDVCDFI VQN NVHRFMMALLDAQHQDICFSACGVLLNLTVDKDKRVILKEGGG
SEG
PRD      ccccccccccccccchhhhhhhhhhhhhc cccceeeecceeeeeeccccccccccc
```

SEQ IKKLVDCRLDGLPTDWQLACLVCKTLWNFSENITNASSCFGNEDTNTLLLLSSFLDEEL  
SEG .....XXXXXXXXXXXXX  
PRD hhhhhhhhccccccchhhhhhhccccccccccccccccceehhhhhhhh

SEQ ALDGSFDPLKNYKHLHWETEFKPVAQQLNRIQRHHTFLEPLPIPSF  
SEG xxx.....  
PRD hhccccccchhhhhhhhhhchhhhhhhhhhhhhheeeccccc

Prosite for DKFZphtes3 14p7.2

PS000001	206->210	ASN_GLYCOSYLATION	PD0C00001
PS000001	212->216	ASN_GLYCOSYLATION	PD0C00001
PS000001	311->315	ASN_GLYCOSYLATION	PD0C00001
PS000001	393->399	ASN_GLYCOSYLATION	PD0C00001
PS000001	493->497	ASN_GLYCOSYLATION	PD0C00001
PS000001	500->504	ASN_GLYCOSYLATION	PD0C00001
PS000001	543->547	ASN_GLYCOSYLATION	PD0C00001
PS000001	584->588	ASN_GLYCOSYLATION	PD0C00001
PS000001	628->632	ASN_GLYCOSYLATION	PD0C00001
PS000001	632->636	ASN_GLYCOSYLATION	PD0C00001
PS000001	635->639	ASN_GLYCOSYLATION	PD0C00001
PS000005	173->176	PKC_PHOSPHO_SITE	PD0C00005
PS000005	186->189	PKC_PHOSPHO_SITE	PD0C00005
PS000005	241->244	PKC_PHOSPHO_SITE	PD0C00005



WO 01/12659

PCT/IB00/01496

PS00005	295->298	PKC_PHOSPHO_SITE	PDOC00005
PS00005	344->347	PKC_PHOSPHO_SITE	PDOC00005
PS00005	387->390	PKC_PHOSPHO_SITE	PDOC00005
PS00005	421->424	PKC_PHOSPHO_SITE	PDOC00005
PS00006	79->83	CK2_PHOSPHO_SITE	PDOC00006
PS00006	201->205	CK2_PHOSPHO_SITE	PDOC00006
PS00006	214->218	CK2_PHOSPHO_SITE	PDOC00006
PS00006	218->222	CK2_PHOSPHO_SITE	PDOC00006
PS00006	230->234	CK2_PHOSPHO_SITE	PDOC00006
PS00006	320->324	CK2_PHOSPHO_SITE	PDOC00006
PS00006	344->348	CK2_PHOSPHO_SITE	PDOC00006
PS00006	439->443	CK2_PHOSPHO_SITE	PDOC00006
PS00006	477->481	CK2_PHOSPHO_SITE	PDOC00006
PS00006	483->487	CK2_PHOSPHO_SITE	PDOC00006
PS00006	654->658	CK2_PHOSPHO_SITE	PDOC00006
PS00006	698->702	CK2_PHOSPHO_SITE	PDOC00006
PS00008	17->23	MYRISTYL	PDOC00008
PS00008	64->70	MYRISTYL	PDOC00008
PS00008	144->150	MYRISTYL	PDOC00008
PS00008	384->390	MYRISTYL	PDOC00008
PS00008	402->408	MYRISTYL	PDOC00008
PS00008	473->479	MYRISTYL	PDOC00008
PS00008	533->539	MYRISTYL	PDOC00008
PS00008	580->586	MYRISTYL	PDOC00008
PS00008	641->647	MYRISTYL	PDOC00008
PS00009	67->71	AMIDATION	PDOC00009

(No Pfam data available for DKF2phtes3\_14p7.2)

DKFZphtes3\_15a13

group: testes derived

DKFZphtes3\_15a13 encodes a novel 387 amino acid protein with weak similarity to *S.cerevisiae* Hop1.

No informative BLAST results; No predictive prosite, pfam or SCOP motif.

The new protein can find application in studying the expression profile of testis-specific genes.

similarity to *S.cerevisiae* Hop1

complete cDNA, complete cds, potential start codon at Bp 116, 3 EST

hits

*S.cerevisiae* Hop1p is a meiosis-specific protein

Sequenced by GBF

Locus: unknown

Insert length: 1848 bp

Poly A stretch at pos. 1766, no polyadenylation signal found

```
1  GGAAAGCGCA  TCGCGGTCGG  GCACAGCGCG  TGCAGCCTCG  TGCAGCTCTT
51  CTGGTCTCCG  GCGCCCGCCC  CTCAGACGTA  ATGTTGAATT  AAAGAAAATA
101  CTTTATCAGA  AGAAGATGGC  CACTGCCGAG  TTGCAGAGGA  CTCCCATGAG
151  TGCACCTGGT  TTCCCAATA  AGATATCAAC  TGAACACCAG  TCTTTGGTGT
201  TAGTGAAGAG  GCTTCTAGCA  GTTTCAGTAT  CCTGTATCAC  GTATTGAGG
251  GGAATATTCC  CAGAATGCGC  TTATGGAACA  AGATATCTAG  ATGATCTTTG
301  TGTCAAAATA  CTGAGAGAA  ATAAAAATTG  CCCAGGATCT  ACACAGTTAG
351  TGAATGGAT  GCTAGGATGT  TATGATGCTT  TACAGAAAAA  ATATGTATAC
401  ACAACCCAG  AAGATCCTCA  GACAATTCCA  GAATGTTACC  AATTCAAATT
451  CAAATACACC  AATAATGGAC  CACTCATGGA  CTTCATAAGT  AAAAACCAAA
501  GCAACGAATC  TAGCATGTG  TCTACTGACA  CCAAGAAAGC  AAGCATTCTC
551  CTCATTGCGA  AGATTTATAT  CCTAATGCAA  AATCTGGGCG  CTTTACCTAA
601  TGATGTTTGT  TTGACCATGA  AACTTTTTTA  CTATGATGAA  GTTACACCCC
651  CAGATTACCA  GCCTCCCGGT  TTTAAGGATG  GTGATTGTGA  AGGAGTTATA
701  TTTGAAGGGG  AACCTATGTA  TTTAAATGTG  GGAGAAGTCT  CAACACCTTT
751  TCACATCTTC  AAAGTAAAAG  TGACCACTGA  GAGAGAACGA  ATGGAAAAA
801  TTGACTCAAC  TATACTATCA  CCAAAACAAA  TAAAAACACC  ATTTCAAAAA
851  ATCTGAGGG  ACAAGATGT  AGAAGATGAA  CAGGAGCATT  ATACAAGTGA
901  TGATTGGAG  ATTGAACATA  AAATGGAAGA  ACAGGAAAAA  AACCCGTCAT
951  CTTCTGAAC  TGAAGAACCA  AGTTTAGTTT  GTGAGGAAGA  TGAAATTATG
1001  AGGTCTAAAG  AAAATCCAGA  TCTTTCTATT  TCTCATTCTC  AGGTTGAGCA
1051  GTTAGTCAAT  AAAACATCTG  AACTTGATAT  GTCTGAAAGC  AAAACAAGAA
1101  GTGGAAAAGT  CTTTCAGAA  AAAATGGCAA  ATGGAAATCA  ACCAGTAAAA
1151  TCTTCCAAAG  AAAATCGGAA  GAGAAGTCAA  CATGAATCTG  GGAGAATAGT
1201  CCTCCATCAC  TTTGATTCTT  CTAGTCAAGA  GTCAAGTCCA  AAAAGGAGAA
1251  AGTTTAGTGA  ACCAAAGGAA  CATATATAAA  AATTATTTTT  GTTCTGCAGG
1301  CTGCGAGAGT  TCTTCTCACC  ATTTAAACTG  AAGGACCCCT  TATTATATTT
1351  CCCTAACTCT  GAAGATGTAT  ATGTAGTTTA  AAGCAGTTTG  TACATAAAA
1401  CTAAGTTTTT  GGCTGACTGT  CATATTGTGG  TCCTTAATCT  TGAGATAAAT
1451  CCAATAGAAC  TTTTGAATA  AAGCAAAAGT  ACAAATGTCA  TAATTGATTC
1501  GGTAAATAAG  AAAATTTCAA  AATTGATTTT  GTTCATTACC  TACTTAATAT
1551  TTCTTTTAAA  TATATACTAA  CTGTAAAGGC  CCTCTAATGC  CATTTTTCTA
1601  AACAGTAATG  TTTACTTTGG  TATTAAATTT  TGGTATGGAT  TCACTTTTTA
1651  CTTATGTTAA  AATTATACCA  TTTAACTGGC  TCTTTGTGCA  TTGTGCTGTT
1701  ATTAACACAA  TGTCTTCAA  TATTTTGACA  TAATGTATTA  ACATTTTAAT
1751  ATATAATGTA  CAATTTAAAA  AAAAAAAAAA  AAAAAAAAAA  AAAAAAAGG
1801  GGCGGCGCT  CTAGAGGATC  CAAGCTTACG  TACAAAAAAA  AAAAAAAGG
```

## BLAST Results

No BLAST result

## Medline entries

No Medline entry

## Peptide information for frame 2

ORF from 116 bp to 1276 bp; peptide length: 387  
Category: similarity to known protein

```
1 MATAQLQRTP MSALVFPNKI STEHQSLVLV KRLLAHSVSC ITYLRGIFPE
51 CAYGTRYLDD LCVKILREDK NCPGSTQLVK WMLGCDALQ KRYVYTNPED
101 PQTISECYQF KFKYTNNGLP MDFISKQSN ESSMLSTDK KASILLIRKI
151 YILMQLGLPL PNDVCLTMKL FYYDEVTPPD YQPPGFKDGD CEGVIFEGEP
201 MYLNVGEVST PFHIFKVKVT TERERMENID STILSPKQIK TPFQKILRDK
251 DVEDEQEHYT SDDLDIETKM EEQEKNPASS ELEEPSLVCE EDEIMRSKES
301 PDLISHSQV EQLVNKTSLE DMSKSTRSG KVFQNMANG NQPVKSSKEN
351 RKRSQHESGR IVLHHFDSSS QESVPKRRKF SEPKEHI
```

## BLASTP hits

No BLASTP hits available

## Alert BLASTP hits for DKF2phtes3\_15a13, frame 2

TREMBL:ATAC2130\_3 product: "F1N21.3"; The sequence of BAC F1N21 from Arabidopsis thaliana chromosome 1, complete sequence., N = 1, Score = 274, P = 5.7e-22

TREMBL:SC9877\_9 gene: "hop1"; S.cerevisiae chromosome IX cosmid 9877., N = 2, Score = 126, P = 7.1e-09

PIR:A34691 meiosis-specific protein HOP1 - yeast (Saccharomyces cerevisiae), N = 2, Score = 126, P = 7.8e-08

>TREMBL:ATAC2130\_3 product: "F1N21.3"; The sequence of BAC F1N21 from Arabidopsis thaliana chromosome 1, complete sequence.  
Length = 562

## HSPs:

Score = 274 (41.1 bits), Expect = 5.7e-22, P = 5.7e-22  
Identities = 84/290 (28%), Positives = 145/290 (50%)

```
Query: 22 TEHQSLVLVKRLLAHSVSCITYLRGIFPECAYGTRYLDDLCVKILREDKNCPGSTQLVKW 81
TE SL+L + LL +++ I+Y+RG+FPE + + + L +KI + S +L+ W
Sbjct: 11 TEQDSLLLRNLLRIAIFNISYIRGLFPEKYFNKSVPALDMKIKRLMPDAESRRLIDW 70

Query: 82 M-LGCDYDALQKKYVYT-----NPEDPQTISECYQFKFYTNNGP--LMDFISK--NQSN 130
M G YDALQ+KY+ T D I E Y F F Y+++ +M I++ N+ N
Sbjct: 71 MKGQVYDALQRKYLKTLMFISICETVDGPMIEE-YSFYSYSDSDSQDVMNINRTGNKKN 129

Query: 131 ESSMLST-----DTKKASILLIRKIYILMQLGLPLNDVCLTMKLFYYDEVTPPDYQPP 184
ST + ++ ++R + LM+ L +P++ + MKL YYD+VTPPDY+PP
Sbjct: 130 GGIFNSTADITPNQMRSSACKMVRTLVLQMLRTLDKMPDERTIVMKLLYYDDVTPPDYEP 189

Query: 185 GFKD--GDCEGVIFEGEPMYLNVGEVSTPFHIFKVKVT-----ERERMENIDSTILS 235
F+ D ++ P+ + +G V++ + +KV + E + M++ D +
Sbjct: 190 FFRGCTEAEQYVWTKNPLRMEIGNVNSKHLVLTLLKVKSVLDPCEDEDDMDQD-DGKSIG 248

Query: 236 PKQIKTPFKILRDKDVEDEQEHY-----TSDDLDIETKMEEQEKNPASSE 281
P + Q D ++ QE+ DD D E ++ +PA +E
Sbjct: 249 PDSVHDD-QPSDSDSEISQTQENQFIVAPVEKQDDDDGEVDEDDNTQDPAENE 300
```

## Pedant information for DKF2phtes3\_15a13, frame 2

## Report for DKF2phtes3\_15a13.2

```
[LENGTH] 387
[MW] 44417.64
[pI] 5.57
[HOMOL] TREMBL:ATAC2130_3 product: "F1N21.3"; The sequence of BAC F1N21 from
Arabidopsis thaliana chromosome 1, complete sequence. 9e-23
[FUNCAT] 09.13 biogenesis of chromosome structure [S. cerevisiae, Y1L072w] 7e-11
[FUNCAT] 03.19 recombination and dna repair [S. cerevisiae, Y1L072w] 7e-11
[FUNCAT] 03.13 meiosis [S. cerevisiae, Y1L072w] 7e-11
[FUNCAT] 30.10 nuclear organization [S. cerevisiae, Y1L072w] 7e-11
[PIRKW] nucleus 2e-09
[PIRKW] zinc finger 2e-09
```

```

[PIRKW]      DNA binding 2e-09
[PROSITE]    MYRISTYL      1
[PROSITE]    CAMP_PHOSPHO_SITE      3
[PROSITE]    CK2_PHOSPHO_SITE      12
[PROSITE]    PKC_PHOSPHO_SITE      7
[PROSITE]    ASN_GLYCOSYLATION      3
[KW]         Alpha_Beta

```

```

SEQ  MATAQLQRTFMSALVFFNKISTEHQSLVLVKRLAVSVSCITYLRGIFPECAYGTRYLDD
PRD  cccccccccccccccccccccccccccccccccccccccccccccccccccccccccchh
SEQ  LCVKILREDKNCPGSTQLVKNMGLCYDALQKKYVYTNPEDPQTISECYQFKFYTNNGPL
PRD  hhhhhhhccccccccccccccccccccccccccccccccccccccccccccccccccccce
SEQ  MDFISKQSNESMLSTDTKKASILLIRKIYILMQNLGPLPNDVCLTMKLFYYDEVTPPD
PRD  eeeeeeeeecccccccccccccccccccccccccccccccccccccccccccccccccccc
SEQ  YQPPGFKDGDCEGVI FEGEPMYLNVGEVSTPFHIFKVKVTTTEREREMENIDSTILSPKQIK
PRD  cccccccccccccccccccccccccccccccccccccccccccccccccccccccccchh
SEQ  TPFQKILRDKDVEDEQEYHTSDDLQIETKMEEQEKNPASSELEEPSLVCEDEIMASKES
PRD  hhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhcc
SEQ  PDLISHSQVEQLVNKTSELDMSSEKTRSGKVFNKMANGNQPVKSSKENRKRKSQHESGR
PRD  cccccchhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhcc
SEQ  IVLHHFDSSSQESVPKRRKFSEPKHI
PRD  eeeeecccccccccccccccccccccccccccc

```

## Prosites for DKFZphtes3\_15a13.2

PS00001	127->131	ASN_GLYCOSYLATION	PDOC00001
PS00001	130->134	ASN_GLYCOSYLATION	PDOC00001
PS00001	315->319	ASN_GLYCOSYLATION	PDOC00001
PS00004	140->144	CAMP_PHOSPHO_SITE	PDOC00004
PS00004	351->355	CAMP_PHOSPHO_SITE	PDOC00004
PS00004	378->382	CAMP_PHOSPHO_SITE	PDOC00004
PS00005	139->142	PKC_PHOSPHO_SITE	PDOC00005
PS00005	167->170	PKC_PHOSPHO_SITE	PDOC00005
PS00005	221->224	PKC_PHOSPHO_SITE	PDOC00005
PS00005	235->238	PKC_PHOSPHO_SITE	PDOC00005
PS00005	329->332	PKC_PHOSPHO_SITE	PDOC00005
PS00005	346->349	PKC_PHOSPHO_SITE	PDOC00005
PS00005	358->361	PKC_PHOSPHO_SITE	PDOC00005
PS00006	96->100	CK2_PHOSPHO_SITE	PDOC00006
PS00006	103->107	CK2_PHOSPHO_SITE	PDOC00006
PS00006	177->181	CK2_PHOSPHO_SITE	PDOC00006
PS00006	221->225	CK2_PHOSPHO_SITE	PDOC00006
PS00006	260->264	CK2_PHOSPHO_SITE	PDOC00006
PS00006	268->272	CK2_PHOSPHO_SITE	PDOC00006
PS00006	280->284	CK2_PHOSPHO_SITE	PDOC00006
PS00006	308->312	CK2_PHOSPHO_SITE	PDOC00006
PS00006	318->322	CK2_PHOSPHO_SITE	PDOC00006
PS00006	346->350	CK2_PHOSPHO_SITE	PDOC00006
PS00006	354->358	CK2_PHOSPHO_SITE	PDOC00006
PS00006	369->373	CK2_PHOSPHO_SITE	PDOC00006
PS00008	84->90	MYRISTYL	PDOC00008

(No Pfam data available for DKFZphtes3\_15a13.2)

DKFZphtes3\_15c24

-----

group: metabolism

DKFZphtes3\_15c24 encodes a novel 404 amino acid protein with strong similarity to 2-hydroxyacid dehydrogenases.

The novel protein contains a D-isomer specific 2-hydroxyacid dehydrogenases signature. Proteins with such a signature have similar enzymatic activities: D-lactate dehydrogenase (EC 1.1.1.28), catalyzes the reduction of D-lactate to pyruvate. D-glycerate dehydrogenase (EC 1.1.1.29) catalyzes the reduction of hydroxypyruvate to glycerate. 3-phosphoglycerate dehydrogenase (EC 1.1.1.95), catalyzes the oxidation of D-3-phosphoglycerate to 3-phosphohydroxypyruvate. Therefore the novel protein is a new 2-hydroxyacid dehydrogenase.

The new protein can find application in modulation of 2-hydroxyacid dehydrogenases-dependent pathways and as a new enzyme for biotechnologic production processes.

strong similarity to C.elegans T03F1.1

potential start at Bp 55 matches kozak consensus PyCCatGG

Sequenced by GBF

Locus: unknown

Insert length: 1956 bp

Poly A stretch at pos. 1929, polyadenylation signal at pos. 1903

```
1 CGAAGGCGGC GCGAAGGCC CGGGCTGGGA GCGTTGGCGG CCGGAGTCCC
51 AGCCATGGCG GAGTCTGTGG AGCGCTGCA GCAGCGGGTC CAGGAGCTGG
101 AGCGGGAAC TGGCCAGGAG AGGAGTCTGC AGGTCCCAG GAGCGGCGAC
151 GGAGGGGGCG GCCGGGTCCG CATCGAGAAG ATGAGCTCAG AGGTGGTGGA
201 TTCGAATCCC TACAGCCGCT TGATGGCATT GAAACGAATG GGAATGTGAA
251 GCGACTATGA GAAAATCCGT ACCTTTGCCG TAGCAATAGT AGGTGTGGT
301 GGAGTAGGTA GTGTGACTGC TGAATGCTG ACAAGATGTG GCATTGGTAA
351 GTTGCTACTC TTTGATTATG ACAAGGTGGA ACTAGCCAAT ATGAATAGAC
401 TTTTCTTCCA ACCTCATCAA GCAGGATTAA GTAAAGTTCA AGCAGCAGAA
451 CATACTCTGA GGAACATTAA TCCTGATGTT CTTTTGAAG TACACAATA
501 TAATATAACC ACAGTGGAAA ACTTCAACA TTTCATGGAT AGAATAAGTA
551 ATGGTGGGTT AGAAGAAGGA AAACCTGTTG ATCTAGTTCT TAGCTGTGTG
601 GACAATTTTG AAGCTCGAAT GACAATAAAT ACAGCTTGTA ATGAAGTTGG
651 ACAAAACATGG ATGGAATCTG GGGTCAGTGA AAATGCAGTT TCAGGGCATA
701 TACAGCTTAT AATTCCTGGA GAATCTGCTT GTTTTGCCTG TGCTCCACCA
751 CTTGTAGTTG CTGCAAAAT TGAATGAAAAG ACTCTGAAAC GAGAAGGTGT
801 TTGTGCAGCC AGTCTTCCTA CCACATGCGG TGTGGTTGCT GGGATCTTAG
851 TACAAAACGT GTTAAAGTTT CTGTTAAATT TTGGTACTGT TAGTTTTTAC
901 CTGGATACA ATGCAATGCA GGATTTTTTT CCTACTATGT CCATGAAGCC
951 AAATCCTCAG TGTGATGACA GAAATTGCAG GAAGCAGCAG GAGGAATATA
1001 AGAAAAAGGT AGCAGCAGCT CCTAAACAAG AGGTTATACA AGAAGAGGAA
1051 GAGATAATCC ATGAAGATAA TGAATGGGGT ATTGAGCTGG TATCTGAGGT
1101 TTCAGAAGAG GAACTGAAAA ATTTTTCAGG TCCAGTTCCA GACTTACCTG
1151 AAGGAATTAC AGTGGCATA CAAATTCCAA AAAAGCAAGA AGATTCTGTC
1201 ACTGAGTTAA CAGTGGAAGA TTCTGGTGAA AGCTTGGAAG ACCTCATGGC
1251 CAAAATGAAG AATATGTAGA TAATGGACTG GGATATATTG TATTTCTCAT
1301 GTTAAAGCCT CTTCCCTTGA AATTAAAAAA AAATTTTAA TGAATAAACT
1351 TAGGGCAACA TTAATTAAATG TATATTCTTA CCTGAATTGT TATACTTTTT
1401 GAAAACTCTG TGAATTCGCT GTTTCTCCCC GCTCCAACGA AATCATTAAC
1451 TCTCCTAAAA TGTGTTTCAT TCTAGTAAGA AAACCTCAAA GGATATTGTA
1501 GGATATAAAT CTTACTTGAA AACATAGCTG TTGAAATGTT TTGGCCTTTT
1551 GGAGTGGGGG AAGGACAAAT CTGATCCTGT AATCTTTTTC TTCCAGTAA
1601 TCCCTTGTGT CTGTTGCATG AGGACATGGA CAATAAAGTA GTATATGATC
1651 CTCAGATACA GGGAGAAGGA CAAGGCATAC AGCTTATTGA TTAGAGCTGG
1701 CAAGCATCTG CTCATTATGT TTGGAATTGC TTTCTATAAG AAAATGCCCC
1751 ACTACTACTA ACTTGATCAA CAATGAATTC AAAATAGTTA ACCTATGAAA
1801 TAACATCCTC TCAAAATGTT GCTGATGAAG TACAAGTTGA AATGTAGTTA
1851 TTGGAAAAGT CTGTAACTG TGGATCATAT ATATTCAAAG TGAGACAAAG
1901 CAAATAAAAA AGCAGCTATT TTCATGAATA GACAAAAAAA AAAAAAAA
1951 AAAAAG
```

## BLAST Results

-----

No BLAST result

## Medline entries

No Medline entry

## Peptide information for frame 1

ORF from 55 bp to 1266 bp; peptide length: 404  
Category: similarity to unknown protein.  
Classification: Metabolism  
Prosites motifs: D\_2\_HYDROXYACID\_DH\_1 (76-105)

1 MAESVERLQQ RVQELERELA QERSLQVPRS GDGGGGRVRI EKMSSEVVDS  
51 NPYSRLMALK RMGIVSDYEK IRTFAVAIVG VGGVGSVTAE MLTRCGIGKL  
101 LLFDYDKVEL ANMNRLLFFQ HQAGLSKVQA AEHTLRNINP DVLFEVHNIN  
151 ITTVENFQHF MDRISNGGLE EGKPVLDVLS CVDNFEARMT INTACNELGQ  
201 TWMSGVSEN AVSGHIQLII PGESACFACA PPLVVAANID ETKLKREGVC  
251 AASLPTTMGV VAGILVQNVL KFLNFGTVS FYLGYNAMQD FFPTMSMKPN  
301 PQCDNRNCRK QEEYKKKVA ALPKQEVQEE EEEIHEDNE WGIELVSEVS  
351 EELKNFSGP VPDLPEDITV AYTIPEKQED SVTELTVEDS GESLEDLMAR  
401 MKNM

## BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKF2phtes3\_15c24, frame 1

TREMBL:CEUT03F1\_11 gene: "T03F1.1"; Caenorhabditis elegans cosmid  
T03F1., N = 1, Score = 1204, P = 1.9e-122

TREMBL:ATAC98\_3 gene: "YUP8H12.3"; Arabidopsis thaliana chromosome 1  
YAC YUP8H12 complete sequence., N = 1, Score = 733, P = 1.5e-72

PIR:A69319 thiamine biosynthesis protein (thiF) homolog - Archaeoglobus  
fulgidus, N = 1, Score = 218, P = 1.8e-17

TREMBL:AF022796\_4 gene: "moeB"; product: "MoeB"; Staphylococcus  
carnosus molybdenum cofactor biosynthetic gene cluster, complete  
sequence., N = 1, Score = 220, P = 3.7e-16

>TREMBL:CEUT03F1\_11 gene: "T03F1.1"; Caenorhabditis elegans cosmid T03F1.  
Length = 419

## HSPs:

Score = 1204 (180.6 bits), Expect = 1.9e-122, P = 1.9e-122  
Identities = 241/367 (65%), Positives = 293/367 (79%)

Query: 37 RVRIEKMSSEVVDSNPYSRLMALKRMGIVSDYEKIRTFVAIVGVGGVGSVTAEMLTRCG 96  
R +IEK+S+EVVDSNPYSRLMAL+RMGIV++YE+IR VA+VGVGGVGSV AEMLTRCG  
Sbjct: 48 RQKIEKLSAEVVDSNPYSRLMALQRMGIVNEYERIREKTVAIVGVGGVGSVVAEMLTRCG 107

Query: 97 IGKLLLFYDKVELANMNRLLFFQHQAGLSKVQAAEHTLRNINPDVLFVHNINNTTVEN 156  
IGKL+LFYDKVE+ANMNRLLF+Q+QAGLSKV+AA TL ++NPDV EVHN+NITT++N  
Sbjct: 108 IGKLLLFYDKVEIANMNRLLFYQPNQAGLSKVEAARDTLIHVNPDVQIEVHNFNITMDN 167

Query: 157 FQHFMDRISNGGLEEGKPVLDVLSVDNFEARMTINTACNELGQTMESGVSENAVSGHI 216  
F F++RI G L +GK +DLVLSVDNFEARM +N ACNE Q WMESGVSENAVSGHI  
Sbjct: 168 FDTFVNRIKGSLLTDGK-IDLVLSVDNFEARMVNMACEENQIMESGVSENAVSGHI 226

Query: 217 QLIIPGESACFACAPPLVVAANIDEKTLKREGVCAASLPTTMGVVAGILVQNVLFLLNF 276  
Q I PG++ACFAC PPLVVA+ IDE+TLKR+GVCAASLPTTM VVAG LV N LK+LLNF  
Sbjct: 227 QYIEPGKTACFACVPLVVASGIDERTLKRKGVCAASLPTTMVAVAGFLVMNTLYLLNF 286

Query: 277 GTVSFYLGYNAMQDFFPTMSMKPNPQCDNRNCRKQEEYKKKVAALPKQ-EV-IQEEEEEI 334  
G VS Y+GYNA+ DFFP S+KPNP CDD +C ++Q+EY++KVA P EV + EEE +  
Sbjct: 287 GEVSQYVGYNALSDFPRDSIKPNPYCDDSHCLQRQKEYEEKVANQPVLEVEVPEEETV 346

Query: 335 IHEDNEWGIELVSEVSEELKNFSGPVPDLPEDITVAYTIPEKQEDSVTELTVEDSGESL 394  
+HEDNEWGIELV+E SE + S + G+ AY P K+ D+ TEL+ + +  
Sbjct: 347 VHEDNEWGIELVNE-SEPSAEQSSSL--NAGTGLKFAE-PIKR-DAQTELSPAQA--AT 399

Query: 395 EDLMAMKKN 403

Sbjct: 400 HDEFMKSIKD 408

Pedant information for DKFZphtes3\_15c24, frame 1  
-----

Report for DKFZphtes3\_15c24.1

```

[LENGTH]      404
[MW]           44863.36
[pI]           4.79
[HOMOL]        TREMBL:CEUT03F1_11 gene: "T03F1.1"; Caenorhabditis elegans cosmid T03F1. 1e-115

[FUNCAT]      h cofactor metabolism [M. influenzae, HI1449] 2e-08
[FUNCAT]      06.07 protein modification (glycosylation, acylation, myristylation,
palmitoylation, farnesylation and processing) [S. cerevisiae, YDR390c UBA2 - E1-like]
4e-07
[FUNCAT]      04.05.05 mrna processing (5'-end, 3'-end processing and mrna degradation) [S.
cerevisiae, YDR390c UBA2 - E1-like] 4e-07
[FUNCAT]      06.13.01 cytoplasmic degradation [S. cerevisiae, YDR390c UBA2 - E1-like]
4e-07
[FUNCAT]      30.10 nuclear organization [S. cerevisiae, YDR390c UBA2 - E1-like] 4e-07
[FUNCAT]      11.01 stress response [S. cerevisiae, YKL210w UBA1 - E1-like] 2e-06
[FUNCAT]      30.03 organization of cytoplasm [S. cerevisiae, YKL210w UBA1 - E1-like]
2e-06
[BLOCKS]      BL01042A Homoserine dehydrogenase proteins
[PIRKW]       thiamine pyrophosphate 1e-07
[PIRKW]       molybdenum 5e-07
[PIRKW]       molybdopterin biosynthesis 5e-07
[SUPPAME]     molybdopterin biosynthesis protein moeB 2e-12
[PROSITE]     D_2_HYDROXYACID_DH_1_1
[KW]          TRANSMEMBRANE 1
[KW]          LOW COMPLEXITY 8.66 %

```

[illegible]

Prosites for DKFZphtes3\_15c24.1

PS00065 76->105 D\_2\_HYDROXYACID\_DH\_1 PDOC00063

(No Pfam data available for DKFZphtes3\_15c24.1)

DKF2phtes3\_15c6

group: transmembrane protein

DKF2phtes3\_15c6 encodes a novel 118 amino acid protein without similarity to known proteins.

The novel protein contains 1 transmembrane region.  
No informative BLAST results; No predictive prosite, pfam or SCOP motif.

The new protein can find application in studying the expression profile of testis-specific genes and as a new marker for testicular cells.

unknown

complete cDNA, complete cds, EST hits

Sequenced by GBF

Locus: unknown

Insert length: 1283 bp

Poly A stretch at pos. 1264, no polyadenylation signal found

```
1 GAGACACTGA GCCCCGAGAC AGTGAGTGGT GGCCTCACTG CTCTGCCCGG
51 CACCCGTGCA CCTCCACTTT GCCTTGTGG AAGTGACCCA GCCCCCTCCC
101 CTTCATTCT CCCACCTGTT CCCCAGGACT CACCCAGCC CCTGCCTGCC
151 CCTGAGGAAG AAGAGGCACT CACCACTGAG GACTTTGAGT TGCTGGATCA
201 GGGGAGAGTG GAGCAGCTGA ATGCAGAGCT GGGCTTGGAG CCAGAGACAC
251 CGCCAAAACC CCCTGATGCT CCACCCCTGG GGGCCGACAT CCATTCTCTG
301 GTACAGTCAG ACCAAGAAGC TCAGGCCCTG GCAGAGCCAT GAGCCAGCCG
351 TTGAGGAAGG AGCTGCAGGC ACAGTAGGGC TTCCTGGCTA GGAGTGTTGC
401 TGTTCCTCC TTTGCCCTACC ACTCTGGGGT GGGCAGTGT GTGGGGAAGC
451 TGGCTGTGG ATGGTAGCTA TTCCACCCTC TGCCTGCCTG CCTGCCCTGT
501 GTCTTGGGCA TGGTGCAGTA CCTGTGCCCTA GGATTGGTTT TAAATTTGTA
551 AATAATTTTC CATTTGGGTT AGTGGATGTG AACAGGGCTA GGAAGTCCT
601 TCCACAGCC TGCCTTGCC TCCCTGCCCTC ATCTCTATTC TCATTCCACT
651 ATGCCCAAG CCCTGGTGGT CTGGCCCTTT CTTTTCTCTC CTATCCTCAG
701 GGACCTGTGC TGCTCTGCC TCATGTCCCA CTGGTGTGTT TAGTTGAGGC
751 ACTTTATAAT TTTCTCTTG TCTTGTTTC TTTCTGCTT TATTTCCCTG
801 CTGTGTCTCT TCCTTAGCAG CTCACCCCA TCCTTTGCCA GCTCCTCCTA
851 TCCCGTGGG ACTGGCCAAG CTTTAGGGAG GCTCCTGGTC TGGGAAGTAA
901 AGAGTAAACC TGGGGCAGTG GGTGAGGCCA GTAGTTACAC TCTTAGGTCA
951 CTGTAGTCTG TGTAACTTC ACTGCATCCT TGCCCAATTC AGCCCGGCT
1001 TTATGATGAC AGGAGAGCAG GGATCCCGCA GTACATGGCC CCAGCACTGG
1051 AGTTGGTGAG CATGTGCTCT CTCTTGAGAT TAGGAGCTTC CTTACTGCTC
1101 CTCTGGGTGA TCCAAGTGA GTGGGACCCC CTACTAGGGT CAGGAAGTGG
1151 AACTAATCAT CTGTGCAGGT GTTGACTTGA AAAATAAAGT GTTGATTGGC
1201 TAAAAAATAA AAAAAAATAA AAAAAAATAA AAGGGCGGCC GCTCTAGAGG
1251 ATCCAAGCTT ACCTAAAAAA AAAAAAATAA AAG
```

## BLAST Results

No BLAST result

## Medline entries

No Medline entry

## Peptide information for frame 2

ORF from 461 bp to 814 bp; peptide length: 118  
Category: putative protein

```
1 MVAIPPSACL PACCPGHGAV FVPRIGFKFV NNFFFLVDV NRAREVLPTA
51 CACLASSSLF SFHYAPSPGG LALSFSSYPQ GPVLLCPHVP LGCLVEALYN
101 FSLVLCSELL YFPAVSCP
```



No BLASTP hits available

PIR:S54250 ribosomal protein L2 - Arabidopsis thaliana, N = 1, Score = 76, P = 0.33

**HSPs:**

```

Query:      15 PGHGAVPVPRIGKFVNNFFPGLVDVNNRAREVLPTACACLPASSLFSFHYAPSPGGLALS 74
             PG GA P+ R+ F+ PF + +E+ A C P SSL+ A G L
Sbjct:      52 PGRGA-PLAGRVTFRH----PF F--KKQKELVAEAECTPVSSLYCGKKATLVGVGNLP 103

Query:      75 FSSYPOGLRVLCP---HV-PLGCLVEALYNFSLVL 105
             S P+G V+ C HV G L A ++++V+
Sbjct:      104 LRSIPEGAVV-CNVEHHVGDGRGVLAARASGDYAIIV 137

```

Report for DKF2phtes3 15c6.2

```
SEQ      MVAIPPSACIAPCCPGHGAVPVPRIGIFKVNFPFGLVDVNRAREVLPTACACLPASSLF
PRD      cccccccccccccccccccccccceeeccccceehhhhhhhcccccceccccccc
MEM

SEQ      SFHYAPSPGGLALSSSYPOGVLVLCPHVPLGCLVEALYNFSLVLCFLLYFPAVSCP
PRD      eeeccccccccceeeccccccccccccccccchhhhhhhhhchhhhhhhhhcccccccc
MEM      .MMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMM
```

PS00001	100->104	ASN GLYCOSYLATION	PDOC00001
PS00008	70->76	MYRISTYL	PDOC00008
PS00029	84->106	LEUCINE ZIPPER	PDOC00029

(No Pfam data available for DKF2phtes3 15c6.2)

DKFZphtes3\_15g14

-----

group: testes derived

DKFZphtes3\_15g14 encodes a novel 701 amino acid protein with weak similarity to *S. cerevisiae* hypothetical protein YOR243c.  
No informative BLAST results; No predictive prosite, pfam or SCOP motife.

The new protein can find application in studying the expression profile of testis-specific genes.

similarity to YOR243c

complete cDNA, complete cds, potential start codon at Bp 35, EST hits

Sequenced by GBF

Locus: unknown

Insert length: 3495 bp

Poly A stretch at pos. 3462, no polyadenylation signal found

```
1 GCCTTCCACT GAACCGAGGC ACTGTTATAG AAGAATGGAA GAAGATACAG
51 ATTATAGAAT CAGGTTTAGT TCTTTGTGTT TCTTTAATGA TCACGTTGGA
101 TTTCATGGCA CTATAAAAAG CTCACCAAGT GACTTTATTG TTATTGAAAT
151 TGATGAACAG GGACAGTTAG TTAATAAGAC CATCGATGAG CCTATTTTCA
201 AGATTAGTGA AATACAACCT GAGCCAAATA ATTTCCCAA AAAACCAAAA
251 CTAGATCTTC AAAATCTGTC CTTAGAAGAT GGAAGAAACC AAGAAGTTCA
301 TACTTTGATT AAGTACACTG ATGGTGACCA AAATCATCAG TCTGGTTCAG
351 AAAAGGAAGA TACTATCGTT GATGGAACTT CCAATGTGA AGAAAAAGCT
401 GATGTTTTAA GTCCTTTTTT GGATGAAAAA ACTCATGAGT TACTGAATAA
451 TTTTGCCTGT GATGTAAGAG AGAAGTGGCT TTCTAAAAA GAGCTAATTG
501 GACTACCTCC TGAATTCTCA ATAGGCAGAA TCCTTGACAA AAACCAGAGG
551 GCTAGTTTAC ACAGTGCCAT TAGGCAGAAA TTTCCATTTT TAGTAACGTG
601 AGGAAAAAAC AGTGAAATTG TTGTAAACC AAATCTTGAA TATAAAGAAC
651 TTTGTCATTT GGTATCTGAA GAGGAAGCAT TTGACTTTTT TAAATATTTG
701 GATGCARAGA AAGAAAAATC CAAATTTACC TTTAAACCTG ATACAAACAA
751 AGACCACAGA AAAGCTGTCC ACCATTTTGT CAACAAAAAG TTTGAAACCC
801 TTGTGGAAC CAAATCTTTT TCTAAATGA ATTGCATGCG TGGTAATCCG
851 AATGTGGTGG TAACAGTAAG ATTTCCGGAA AAAGCACACA AACGTGGGAA
901 AAGGCCCTCT TCTGAATGCC AAGAAGGAAA AGTTATATAT ACAGCTTTTA
951 CCCTACGAAA GGAACACCTG GAAATGTTTG AAGCGATTGG TTTTITAGCT
1001 ATCAAACTTG GTGTTATTCC TTCGGATTTT AGTTATGCAG GCCTTAAAGA
1051 CAAGAAAGCC ATCACCATATC AAGCAATGGT TGTAGAAAAA GTGACTCCAG
1101 AGAGGTTGAA AAATATTGAA AAAGAAATTG AAAAGAAAAG AATGAATGTC
1151 TTTAATATTC GGTCTGTAGA TGATTCCTCG AGACTTGGTC AGCTCAAAGG
1201 AAATCACTTT GATATTGTCA TTAGAAATTT AAAAAAACAA ATAAATGATT
1251 CTGCAAACTT GAGGGAGAGA ATTATGGAA CAATAGAAAA TGTAAAGAAA
1301 AAAGGCTTTG TGAATTACTA TGGACCACAG AGATTGGGA AGGGAAGGAA
1351 AGTTCACACA GACCAAAATT GACTAGCTTT GCTGAAGAAT GAAATGATGA
1401 AAGCCATAAA ATTGTTTCTT ACACCAGAAG ACTTGGATGA TCCTGTAAT
1451 AGAGCAAAAG AGTATTTTCT TCAAACTGAG GATGCTAAAG GCACACTTTC
1501 ATTGATGCCT GAATTCAAAG TCCGTGAGAG AGCATTTGTT GAGGCATTGC
1551 ACCGCTTTGG CATGACCAGG GAAGGTTGTA TCCAGGCATG GTTCTCTTTA
1601 CCCCATTCCA TCGCATATT CTATGTTTAC GCATATACCA GCAAAATTTG
1651 GAATGAGGCA GTATCTTACA GACTTGAAAC CTATGGAGCA AGAGTAGTGC
1701 AGGGTGATTT GGTCTGTTTG GATGAAGACA TTGATGACGA GAATTTCCCA
1751 AATAGTAAAA TTCACCTGGT AACTGAAGAG GAGGGATCAG CTAATATGTA
1801 TGCAATACAT CAGGTGGTTC TTCCAGTACT TGGATACAAT ATTCACTACC
1851 CGAAGAACAA AGTAGGGCAG TGGTACCATT ACATACTTAG CAGAGATGGA
1901 CTACAGACAT GTAGGTTTAA AGTACCTACT CTGAACTGA ATATACCAGG
1951 TTGCTATAGA CAGATTTTGA AACATCCCTG TAATCTCTCA TACCAACTAA
2001 TGAAGATCA TGACATTGAT GTCAAAACGA AAGGTTCCCA CATTGATGAA
2051 ACAGCTTTGT CTCTTTTGAT CTCTTTTGAT CTTGATGCTT CATGCTATGC
2101 TACCGTTTGT CTGAAGGAAA TAATGAAGCA TGACGTTTAA AACTGATACC
2151 CTTGGTATAA CCATATATAT GTCAACCTTT CTTGTTTTTG AAATTATTGA
2201 TCAGAACAAAT ATACAAGGGA AATGCCATAC CTCTGTTTGT GATAGATACC
2251 CCAGAGTAGT TATTACCTCT TTGTGAGATA AGTAATCTTT GATGAAGATT
2301 GAAATACAAT TTCTCATCCA ATTTTATAT CTGGCATACT GCTGACCTTC
2351 TTGACCACTT GTAATTTTTT CATATTATCT AAAACAGGTG TTAGAGTCAG
2401 ACAGATTCAT TCTTAGATTG TAGCTCTGAC ACTTACTAGT GATTTTGAGT
2451 ATGTTGTTGA TTTTTTTGTG TGTGGTTACT GATAGAATCA AGACAATTAC
2501 AACTTCATAA ATGACAAATA ATAGGATTAT CTCCACATTT TCTGTTGCTG
2551 GAGGAACAAA ACATTGTGCC CATTGAAAAA TTTTAATTTT TGTGTTGTTA
2601 ACTATCCAC ATTATAAATC ATCCTTCACC ATTTATATC AGTTAAATAT
2651 GGGTGTGTTG GGGAGGAATG ACTGCCATGT AGACATGTAT TGATTTAGGA
2701 AGATCTGAGC ATTTCTTTCA TTGTTGGTAA GATATAATGA TGAAATTTAA
```

```
2751 AAAGCAGTAT GGAGCATTAT ATATCAGTAA TGTGATATAT ATACTTAAGC
2801 CAGTTTAAAC ATTTTGGGAA ATGTTAGCAT TAGGAAATAA AATCCAAAAG
2851 AAGGAAGAGA AGCTATATGC AATGCAAAAT TTGCTTATTG CAATATTTTC
2901 ATATACAGAC ACTAAAAACA GTTTTCAAAG TCCAGCATTG CGTAACTAAA
2951 GTAAGTAAAA TGATGTGTAT CAACTTGATG GTAAAAATATG TAGTTATTTA
3001 AAAAAGCAAT GAACAATTTA GTTTCATGAG AAAATGTTGC CCCCTAAAAG
3051 TAGAACACAT ATGTTACAAC TGCAATAATA CTCTGAATTC ATCTTTTACA
3101 AATAAGAGAC ATGTTAGCAT AGTGATTAAT AGCACAGATA TTGGAGACAA
3151 ACTAACCAG TTTGAACCCCT GGCCTGCCA CGTATAGCAC TGCAGCCTTG
3201 GGAAGTTAT TTAACCTCAT GGGCTTCAGT TTCAACATCT GTAAAAATGGG
3251 CATGTTAACA TTGCCTACCT CATAGGATTA CTGTGAGAAT TTTCTAAGTT
3301 AATATATGTA AAGCACTTT AAAAAGTGCC TGGCACTTAG TTATTGTTAA
3351 GTAAGTGTCT GCAGATGCAA GTTTGGAAGA GAAAAGCAAA TAAATGAAAA
3401 TCCCTTCTG TTAAGATGAA AAAAAAAAAA AAAAAAAAAA AAAAAAGGGG
3451 CGGCCGCTCA AGATGAAAAA AAAAAAAAAA AAAAAAAAAA AAAGG
```

## BLAST Results

No BLAST result

## Medline entries

No Medline entry

## Peptide information for frame 2

ORF from 35 bp to 2137 bp; peptide length: 701  
Category: similarity to unknown protein

```
1 MEEDTDYRIR FSSLCFFNDH VGPHGTIKSS PSDFIVIEID EQQQLVNKTI
51 DEPIFKISEI QLEPNNFPKK PKLDLQNLSL EDGRNQEVHT LIKYTDGDDQ
101 HQSGSEKEDT IVDGTSKCEE KADVLSSFLD EKTHELLNMF ACDVREKWL
151 KTELIGLPPE FSGIRILOKN QRASLHSAIR QKFPFLVTVG KNSEIVVKPN
201 LEYKELCHLV SEEEAFDFFK YLDAKKENSK FTFKPDTNKD HRKAVHHFVN
251 KKFGNLVETK SFSKMNCAG NPNVVVTVRF REKAHKGKGR PLSECQEGKV
301 IYTAFTLRKE NLEMFEAIGF LAIKLGVIPS DFSYAGLKDK KAITYQAMVV
351 RKVTPERLKN IEKEIEKKRM NVFNIRSVDD SLRLGQLKGN HFDIVIRNLK
401 KQINDSANLR ERIMEAIENV KKGKGVNYYG PORFGKGRKV HTDQIGLALL
451 KNEMMKAIKL FLTPEDLDDP VNRKKYFLQ TEDAKGTLSL MPEFKVRERA
501 LLEALHFRFGM TEEGCIQAWF SLPHSMRIFY VHAYTSKIWN EAVSYRLETY
551 GARVVQGDV CLDEDDIDEN FPNKSHLVT EEEGSAHMYA IHQVVLVPLG
601 YNIQYPKNKV GQWYHDILSR DGLQTCRFKV PTLKLNIPGC YRQILKHPCN
651 LSYQLMEDHD IDVKTKGSHI DETALSLNIS FDLASCYAT VCLKEIMKHD
701 V
```

## BLASTP hits

No BLASTP hits available

## Alert BLASTP hits for DKF2phtes3\_15g14, frame 2

TREMBL:SPBC1A45P\_10 gene: "SPBC1A4.09"; product: "hypothetical protein"; S.pombe chromosome II cosmid c1A4 left hand region 1-26184 bp  
Originates from chimeric cosmid., N = 3, Score = 511, P = 2.9e-57

PIR:S67136 hypothetical protein YOR243c - yeast (Saccharomyces cerevisiae), N = 2, Score = 516, P = 7.3e-54

SWISSPROT:YQ4B\_CAEEL HYPOTHETICAL 64.6 KD PROTEIN B0024.11 IN CHROMOSOME V., N = 2, Score = 386, P = 2.1e-34

>PIR:S67136 hypothetical protein YOR243c - yeast (Saccharomyces cerevisiae)  
Length = 676

## HSPs:

Score = 516 (77.4 bits), Expect = 7.3e-54, Sum P(2) = 7.3e-54  
Identities = 151/498 (30%), Positives = 245/498 (49%)

Query: 191 KNSEIVKPNLEYKELCHLVSEEEAFDFFK-YLDAKKENSKFTFKPDTNKDHRKAVHHFV 249  
+ E V P L +L + EE+ Y A K + F+ +K R +H +

Sbjct: 109 RRQEFNVDPRLR-NQLVEIFGEEDVLKIESVYRTANKMETAKNFE---DKSVRTKIHQLL 164

Query: 250 NKKFGNLVETKSF SKMNC SAGNPVVTVRFREKAHK-RGKRPLSECQEG-KVIYTAFTL 307  
+ F N +E+ + N +EK ++ R + G + FTL

Sbjct: 165 REAFKNELESVTTDTNTFKIARSNRNRTNKQEKINQTRDANGVENWGYGPKDFIHFTL 224

Query: 308 RKENLEMEFAIGFLAIKLGVIPSD-FSYAGLKDKKAITQAMVVRKVTPERLKNIEKEIE 366  
KEN + EA+ + KL +PS YAG KD++A+T Q + + K+ +RL + + +

Sbjct: 225 HKENKDTMEAVNVIT-KLLRVPSRVIRYAGTKDRRAVTCQRVISIKIGLORLNALNRTL- 282

Query: 367 KKRMMNVFNIRSVDDSLRLGQLKGNHFDIVIRNLKKQINDSANLRERIMEAIENVKKGFV 426  
K M + N D SL LG LGKN F +VIR++ N +L E + +++ + GF+

Sbjct: 283 -KGMIIIGNYNFSDASLNLGDLKGNFVTVIRDVTTG-NSEVSLEEIVSNGCKSLSENGFI 340

Query: 427 NYYGPQRFGKGRKVHTDQIGLALLKNEMMKAIKFLTPEDLDDPVNR-AKKYFLQTEDAK 485  
NY+G QRFG + T IG LL + KA +L L+ +D P ++ A+K + +T+DA

Sbjct: 341 NYFGMQRFQGT-F-SISTHTIGRELLSNWKKAAELILSDQNVLPKSKKARKIWAETKDAA 399

Query: 486 GTLSLMPEFKVRERALLEALHRFGMTEEGCIQ---AWFS----LPHSMRIFYVHAYTSKIW 539  
L MP + E ALL +L E+G A+++ +P ++R YVHAY S +W

Sbjct: 400 LALKQMPROCLAENALLYSLSNQRKEEDGTYSENAYYTAIMKIPRNLRMTVMYHAYQSYVW 459

Query: 540 NEAVSYRLETYGARVVQGDVLC-----LDEDIDDENFPNS-----KIHVTEEEGS 585  
N S R+E +G ++V GDLV L IDDE+F + VT+E+

Sbjct: 460 NIASKRIELHGLKLVVGDVVIDTSEKSPGISDDEDFEDVREAQFIRAKAVTQEDID 519

Query: 586 ANMYAIHQVVLPLVGYNIQYFKNK-VGQWYHDILSRDGLQTCRFKVPVTLKLNIPGCYRQI 644  
+ Y + VVLP G+++ YP N+ + Q Y DIL D + + ++ G YR +

Sbjct: 520 SVKYTMEDVVLPSPGFDVLYPSNEELKQLYVDILKADNMDFNMRKVRDFSLAGSYRTV 579

Query: 645 LKHPCNLSYQLMEDHDIDVKTGSHID 671  
++ P +L Y+++ D + + +D

Sbjct: 580 IQPKSLEYRIIHYDDPSQQLVNTDLD 606

Score = 86 (12.9 bits), Expect = 3.2e-01, Sum P(2) = 2.8e-01  
Identities = 40/160 (25%), Positives = 77/160 (48%)

Query: 22 GFHGTIKSSPSDFVIEIDEGQLVNKTIDEPIFKISEIQLEPNNFPPKPKLDLQNLSE 81  
GF G IK +DF+V EID++G++++ T D+ FK+ + +P K +++ + S E

Sbjct: 55 GFRGQIKQRYTDFLVNEIDQEGKVIHLT-DKG-FKMPK---KPQR--SKEEVNAEKES-E 106

Query: 82 DGRNQEVHTLIKTYDGDQNHQSGS--EKEDTI-VDGTSKCEKADVLSSFLDEKTHELLN 138  
R QE + D + +Q +ED + ++ + K + +F D+ ++

Sbjct: 107 AARRQEFNV----DPELRNQLVEIFGEEDVLKIESVYRTANKMETAKNFEDKSVRTKIH 161

Query: 139 NFACDVREKWLKTELGIPPE-FSIGRILDKNQASLHSAIRQ 181  
+RE + ++ E + F I R ++N R + I Q

Sbjct: 162 QL---LREAFKNELESVTTDTNTFKIARS-NRNSRTNKQEKINQ 201

Score = 58 (8.7 bits), Expect = 7.3e-54, Sum P(2) = 7.3e-54  
Identities = 10/23 (43%), Positives = 17/23 (73%)

Query: 676 SLLISFDLDASCYATVCLKEIMK 698  
++++ F L S YAT+ L+E+MK

Sbjct: 638 AVVLKFLQLGTSAYATMALRELK 660

Pedant information for DKF2phtes3\_15g14, frame 2

Report for DKF2phtes3\_15g14.2

[LENGTH] 701  
[MW] 80700.96  
[pI] 7.31  
[HOMOL] PIR:S67136 hypothetical protein YOR243c - yeast (Saccharomyces cerevisiae) 2e-51  
[FUNCAT] 99 unclassified proteins [S. cerevisiae, YOR243c] 8e-53  
[BLOCKS] BL01268C  
[BLOCKS] BL01268B  
[BLOCKS] BL01268A  
[SUPFAM] hypothetical protein HI0701 3e-06  
[PROSITE] MYRISTYL 7  
[PROSITE] AMIDATION 2  
[PROSITE] CAMP\_PHOSPHO\_SITE 1  
[PROSITE] CK2\_PHOSPHO\_SITE 16  
[PROSITE] TYR\_PHOSPHO\_SITE 1  
[PROSITE] PKC\_PHOSPHO\_SITE 13  
[PROSITE] ASN\_GLYCOSYLATION 5  
[KW] Alpha\_Beta

```

SEQ MEEDTDYRIRFSSLCFFNDHVGPHGTIKSSPSDFIVIEIDEQGQLVNKTIDEPIFKISEI
PRD ccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc
SEQ QLEPNFPKKPKLDLQNLSDLEGRNQEVHTLIKYTDGQDNHQSSEKEDTIVDGTSKCEE
PRD cccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccch
SEQ KADVLSSFLDEKTHELLNMFACDVREKWLKTELGPPFESIGRILOKNQRASLSAIR
PRD hhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh
SEQ QKFPFLVTVGKNSEIVVKPNLEYKELCHLVSEEAFFDKYLDACKENSKFTFKPDNTKD
PRD hccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccch
SEQ HRKAVHHFVNKKFGNLVETKSFKMNCSAGNPVVTVRFREKAHKGKRPLSECQEGKV
PRD hhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhcc
SEQ IYTAFTLRKENLEMFAGFLAIKLGVIPSDFSYAGLKDKKAITQAMVVRKVTPERLKN
PRD eeeeeccccccccccccccccccccccccccccccccccccccccccccccccccccccccch
SEQ IEKEIEKKRMNVFNIRSVDDSLRLGQLKGNHFDIVIRNLKQINDSANLRERIMEAIENV
PRD hhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh
SEQ KKGKFVNYYGQRFQGRKRVHTDQIGLALLKNEMMKAIKFLTPEDLDDPVNRKAYFLQ
PRD hccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccch
SEQ TEDAKGTLMLPEFKVRERALLEALHRFGMTEEGCIQAWFSLPHSMRI FYVHAYTSKIWN
PRD hccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccch
SEQ EAVSYRLETYGARVVQGDVLCLEDDIDDENFPNSKIHLVTEEGSANMYAIHQVVLVPLG
PRD hhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhcc
SEQ YNIQYPKNKVGQWYHDILSRDGLQTCRFKVPVLKLNIPGCYRQILKHPCNLSYQLMEDHD
PRD cccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccch
SEQ IDVKTGSHIDETALLISFDLDASCYATVCLKEIMKHDV
PRD cccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccch

```

## Prosites for DKFZphtes3\_15g14.2

PS00001	47->51	ASN_GLYCOSYLATION	PDOC00001
PS00001	77->81	ASN_GLYCOSYLATION	PDOC00001
PS00001	266->270	ASN_GLYCOSYLATION	PDOC00001
PS00001	404->408	ASN_GLYCOSYLATION	PDOC00001
PS00001	650->654	ASN_GLYCOSYLATION	PDOC00001
PS00004	351->355	CAMP_PHOSPHO_SITE	PDOC00004
PS00005	26->29	PKC_PHOSPHO_SITE	PDOC00005
PS00005	105->108	PKC_PHOSPHO_SITE	PDOC00005
PS00005	115->118	PKC_PHOSPHO_SITE	PDOC00005
PS00005	232->235	PKC_PHOSPHO_SITE	PDOC00005
PS00005	237->240	PKC_PHOSPHO_SITE	PDOC00005
PS00005	277->280	PKC_PHOSPHO_SITE	PDOC00005
PS00005	306->309	PKC_PHOSPHO_SITE	PDOC00005
PS00005	381->384	PKC_PHOSPHO_SITE	PDOC00005
PS00005	525->528	PKC_PHOSPHO_SITE	PDOC00005
PS00005	535->538	PKC_PHOSPHO_SITE	PDOC00005
PS00005	544->547	PKC_PHOSPHO_SITE	PDOC00005
PS00005	625->628	PKC_PHOSPHO_SITE	PDOC00005
PS00005	632->635	PKC_PHOSPHO_SITE	PDOC00005
PS00006	30->34	CK2_PHOSPHO_SITE	PDOC00006
PS00006	49->53	CK2_PHOSPHO_SITE	PDOC00006
PS00006	79->83	CK2_PHOSPHO_SITE	PDOC00006
PS00006	95->99	CK2_PHOSPHO_SITE	PDOC00006
PS00006	103->107	CK2_PHOSPHO_SITE	PDOC00006
PS00006	105->109	CK2_PHOSPHO_SITE	PDOC00006
PS00006	110->114	CK2_PHOSPHO_SITE	PDOC00006
PS00006	116->120	CK2_PHOSPHO_SITE	PDOC00006
PS00006	127->131	CK2_PHOSPHO_SITE	PDOC00006
PS00006	150->154	CK2_PHOSPHO_SITE	PDOC00006
PS00006	211->215	CK2_PHOSPHO_SITE	PDOC00006
PS00006	237->241	CK2_PHOSPHO_SITE	PDOC00006
PS00006	377->381	CK2_PHOSPHO_SITE	PDOC00006
PS00006	463->467	CK2_PHOSPHO_SITE	PDOC00006
PS00006	580->584	CK2_PHOSPHO_SITE	PDOC00006
PS00006	668->672	CK2_PHOSPHO_SITE	PDOC00006
PS00007	537->546	TYR_PHOSPHO_SITE	PDOC00007
PS00008	25->31	MYRISTYL	PDOC00008
PS00008	43->49	MYRISTYL	PDOC00008
PS00008	114->120	MYRISTYL	PDOC00008

WO 01/12659

PCT/IB00/01496

PS00008	326->332	MYRISTYL	PDOC00008
PS00008	385->391	MYRISTYL	PDOC00008
PS00008	514->520	MYRISTYL	PDOC00008
PS00008	622->628	MYRISTYL	PDOC00008
PS00009	287->291	AMIDATION	PDOC00009
PS00009	436->440	AMIDATION	PDOC00009

(No Pfam data available for DKFZphes3\_15g14.2)

DKFZphtes3\_15h1

-----

group: testes derived

DKFZphtes3\_15h1 encodes a novel 672 amino acid protein with very weak similarity to several proteins.

No informative BLAST results; No predictive prosite, pfam or SCOP motif.

The new protein can find application in studying the expression profile of testis-specific genes.

similarity to Hsp70/Hsp90 organizing protein

complete cDNA, complete cds, no EST hits

Sequenced by GBF

Locus: unknown

Insert length: 2277 bp

Poly A stretch at pos. 2252, polyadenylation signal at pos. 2226

```
1 AAACAGATA GAGGTTCTCC AGCTTTTCTT TGATTGTCTC TGCTTTAGCG
51 TCTCTAAATC CGGTCAACAT GTCGGACCCC GAAGGCGAGA CCTTGCGAAG
101 CACCTTTCCC TCTTATATGG CCGAAGCGGA GCGGCTCTAC CTGTGCGGGG
151 AATTTTCTAA AGCCGCGCAG AGCTTCAGCA ACGCTCTTTA CCTTCAGGAT
201 GGAGACAAGA ACTGCCTGGT TGCTCGCTCA AAGTGCTTCC TGAAGATGGG
251 AGACTTGGAG AGATCCCTGA AGGATGCTGA GGCTTCGCTC CAGAGTGACC
301 CAGCTTTCTG TAAGGGGATT TTGCAAAAGG CTGAGACACT GTACACCATG
351 GGAGACTTTG AGTTTGCCCTT GGTATTCTAT CATCGAGGCT ACAAGCTGAG
401 GCCTGATCGG GAATTCAGAG TTGGCATTCA GAAAGCCGAG GAAAGCATCA
451 ACAACTCAGT GGAAGTCCT TCTTCATTA AGCTGGAGAA CAAAGGGGAC
501 CTCTCCTTCT TAAGCAAGCA GGCTGAGAAT ATAAAAGCCC AGCAGAAGCC
551 TCAGCCCATG AACACCTCT TACACCCAC CAAGGGAGAG CCCAAGTGGA
601 AGGCCCTCGT CAAGAGTGAG AAGACTGTCC GCCAGCTTCT GGGGAGCTC
651 TACGTGGACA AAGAGTATTT GGAGAAGCTC CTATTGGATG AAGACCTGAT
701 CAAAGGCACC ATGAAGGCGC GCCTGACTGT GGAGGACCTC ATCATGACCG
751 GCATCAACTA CCTGGATACT CACAGCAACT TCTGGAGGCA GCAGAAGCCG
801 ATCTACGCCA GGGAGCGGGA CCGGAAGCTG ATGCAAGAGA AATGGCTCGG
851 GGACCACAAA CGCCGTCCTT CACAGACAGC CCATTACATC CTCGAAGGCC
901 TGGAGGACAT TGATATGTTG CTCACAAGTG GCAGTGCTGA AGGGAGCTTT
951 CAGAAAGCTG AGAAAGTGCT GAAGAAGGTA CTGGAATGGA ACAAGGAAGA
1001 GGTACCCAAC AAGGATGAAC TGGTTGGAAA CTGTATAGC TGCAATGGGA
1051 ATGCCAGAT TGAGCTGGGG CAGATGGAGG CAGCCCTGCA GAGCCACAGA
1101 AAGGACCTGG AGATCGCCAA GGAATATGAC CTTCCTGATG CAAAATCGAG
1151 AGCCCTTGAC AACATTGGCA GAGTTTTTGC CAGAGTTGGG AAATTCCAGC
1201 AAGCCATTGA CACGTGGGAA GAAAGATCC CTCTGGCAAA AACCACCTCG
1251 GAGAAGACCT GGCTGTTCCA CGAGATCGGC CGCTGCTACT TGGAGCTGGA
1301 CCAGGCTCGG CAGGCCCAGA ATTATGGCGA GAAGTCCAG CAGTGTGCCG
1351 AGGAGGAAGG GGACATTGAG TGGCAACTGA ATGCCAGTGT TCTGGTGGCC
1401 CAGGACCAAG TGAAGCTGAG AGACTTCGAG TCAGCCGTGA ACAATTTTGA
1451 GAAGGCCCTG GAGAGAGCAA AGCTTGTGCA TAACAACGAG GCGCAGCAGG
1501 CCATCATCAG TGCCTTGGAC GATGCCAACA AGGATATCAT CAGAGAAGCTG
1551 AGGAAAACCA ACTACGTGGA GAATCTCAAA GAAAAAAGCG AGGGAGAAGC
1601 TTCACTGTAT GAAGATAGAA TAATAACAAG AGAGAAGGAC ATGAGGAGAG
1651 TGAGAGATGA GCCCAGAGAAG GTGCTGAAGC AGTGGGACCA TAGTGAGGAT
1701 GAGAAAGAGA CAGATGAGGA CGATGAGGCT TTTGGGGAAG CTCTGCAGAG
1751 CCCAGCAAGC GGAAGCAGA GTGTGGAAGC AGGAAAAGCC AGAAGCGATT
1801 TGGGAGCAGT TGCCAAGGGC CTGTGAGGAG AATTAGGCAC AAGATCAGGA
1851 GAAACAGGCA GGAAGCTACT AGAAGCTGGC AGAAGAGAGT CAAGAGAAAT
1901 TTATAGGAGG CCTTCGGGAG AATTAGAGCA AAGACTCTCA GGAGAATTCA
1951 GCAGACAGGA ACCAGAAGAA CTAAAGAAAC TTTAGAAAGT GGGCAGAAAG
2001 GAGCCAGAAG AACTGGGAAA AACACAATTT GGAGAAATAG GAGAAACGAA
2051 AAAAACAGGA AATGAGATGG AAAAGGAATA TGAATGAAGC CATCGGTAGA
2101 GATGAGGATC AGGAAGCTGG TGTTCAGAGG GATCATGGGA TTTTATTAAA
2151 CTGGATTTTC AAGCGATTGG TCTGTTATAG GAAAAATGAG GGTTTTACTT
2201 CTGCTGCTTT CCATCACTAT TTTGCCATTA AATAGGTGTC TTTCACCTTT
2251 GCAAAAAATA AAAAAAATA AAAAAA
```

## BLAST Results

-----

No BLAST result

## Medline entries

No Medline entry

## Peptide information for frame 3

ORF from 69 bp to 2084 bp; peptide length: 672  
Category: similarity to known protein

1 MSDPEGETLR STFFSYMAEG ERLYLCGEFS KAAQSFSNAL YLQDGDKNCL  
51 VARSKCFLKM GDLSRLKDA EASLQSDPAF CKGILQKAET LYTMGDFEFA  
101 LVFYHRGYKL RPDREFRVGI QKAQEAINNVS VGSPSSIKLE NKGDLNFLSK  
151 QAENIKAAQOK PQPMKHLHP TKGEPEKWKAS LKSEKTVRQL LGELYVDKEY  
201 LEKLLDEDL IKGTMKGGTL VEDLIMTGIN YLDTSHNFWR QOKPIYARER  
251 DRKLMQEKWL RDHKRRPSQT AHYILKSLED IDMLLTSGSA EGSLOKAERV  
301 LKKVLEWNEK EVPNKDELVG NLYSCIGNAQ IELQMEAAAL QSHRKDLLEIA  
351 KEYDLPOAKS RALDNIGRVF ARVGKFOQAI DTWEEKIPLA KTTLEKTWLF  
401 HEIGRCYLEL DQAWQAQNYG EKSQCAEEEE GDIEWQLNAS VLVAQAQVKL  
451 RDFSASVNNF EKALERAQKV HNNEAQAII SALDDANKGI IRELKNTYV  
501 ENLKEKSEGE ASLYEDRIIT REKDMRRVRD EPEKVVQWMD HSEDEKETDE  
551 DDEAFGEALQ SPASGKQSV EAGKARSDLGA VAKGLSGELG TRSGETGRKL  
601 LEAGRESRE IYRRPSGELE QRLSGEFSRQ EPEELKKLSE VGRREPEELG  
651 KTFGEIGET KKTGNEMEKE YE

## BLASTP hits

Entry AF039202.1 from database TREMBL:  
product: "Hsp70/Hsp90 organizing protein"; *Cricetulus griseus*  
Hsp70/Hsp90 organizing protein mRNA, complete cds.  
Score = 149, P = 5.3e-07, identities = 42/160, positives = 74/160

Entry A109782.1 from database TREMBL:  
product: "myosin heavy chain"; *Argopecten irradians* myosin heavy chain  
mRNA, complete cds.  
Score = 155, P = 6.1e-07, identities = 140/623, positives = 256/623

Entry S56658 from database PIR:  
stress-induced protein stil - soybean  
Score = 156, P = 9.7e-08, identities = 41/153, positives = 72/153

Alert BLASTP hits for DKFZphtes3\_15h1, frame 3

No Alert BLASTP hits found

Pedant information for DKFZphtes3\_15h1, frame 3

## Report for DKFZphtes3\_15h1.3

[LENGTH] 672  
[MW] 76655.61  
[pI] 5.49  
[HOMOL] PIR:S56658 stress-induced protein stil - soybean 6e-10  
[SUPFAM] tetratricopeptide repeat homology 1e-07  
[PROSITE] MYRISTYL 7  
[PROSITE] AMIDATION 3  
[PROSITE] CAMP\_PHOSPHO\_SITE 4  
[PROSITE] CK2\_PHOSPHO\_SITE 15  
[PROSITE] TYR\_PHOSPHO\_SITE 1  
[PROSITE] PKC\_PHOSPHO\_SITE 11  
[PROSITE] ASN\_GLYCOSYLATION 2  
[KW] All\_Alpha  
[KW] LOW\_COMPLEXITY 4.76 %

SEQ MSDPEGETLRSTFFSYMAEGERLYLCGEFSKAAQSFSNALYLQDGDKNCLVARSKCFLKM  
SEG .....  
PRD cccccccceccccccccccccccccchhhhhhhhhhhhhccccceehhhhhhhhhhh  
  
SEQ GDLSRLKDAEASLQSDPAFCCKGILQKAETLYTMGDFEFALVYFYHRGYKL RPDREFRVGI  
SEG .....  
PRD hcchhhhhhhhhhhccccchhhhhhhhhhhhhhhhhhhhhhhhhhhccccccccchhhhh



```

SEQ      QKAQEAINNVSVPSSIKLEKNGDLSFLSKQAENIAKQKQPQPMKHLHPTKGPEKWKAS
SEG
PRD      hhhhhhhhhhhhhhhhhhhhhhhccchhhhhhhchhhhhhhhhccchhhhhhhccccccccchhh

SEQ      LKSEKTVRQLLGELYVDKEYLEKLDDLEDLIGTKMGGLTVEDLIMTGINYLTHSNFWR
SEG      .....XXXXXXXXXXXXXXXXXXXXX
PRD      hhhhhhhhhhhhhhhhhhhhhhhccchhhhhhhhhhhhhhhhhhhcccccccccccc

SEQ      QQKPIYARERDRKKLQWEKVLROHKRRPSQTAYHILKSLEDIOMLTSGSAEGLQAKAEV
SEG
PRD      cchhhhhhhhhhhhhhhhhhhhhccccccchhhhhhhhhhhheeeccccccchhhhhhhhh

SEQ      LKVLKLEMKEEVPPKDELVGNLYSICGNAQILQGMEEALQSHRKDLIAKEYDLPKAS
SEG
PRD      hhhhhhhccccccccceccccccccchhhhhhhhhhhhhhhhhhhhhhhhhhhccccchh

SEQ      RALDNI GRV FARVKGQQAIDTWEKILPLAKTTLEKTWLFHEIGRCYLELDQAWQAQNYG
SEG
PRD      hhhccccchhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhccccchhhhhhhhhhhhh

SEQ      EKSQQCAEEEGDIEWQLNASVLVAQAQVKLRDFESAVNNFEKALERAKLVHNNEAQAQII
SEG
PRD      hhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhccccchhhh

SEQ      SALDDANKGI IRELKRTNYVENLKEKSGEASLYEDRI I TREKDMRRVRDEPEKVVVKQD
SEG
PRD      hhhccchhhhhhhhhhhhhhhhhhhhhccchhhhhhhhhhhhhhhhhhhhhhhccccceeecc

SEQ      HSEDEKETDEDEAFGEALQSPASGQSV EAGKARSDLGA VAKGLSGELGTRSGETGRKL
SEG      XXXXXXXXXXXXXXXX
PRD      cccccccccchhhhhhhhhccccccccchhhhhhhccccccccceccccccccccccchhh

SEQ      LEAGRRRESREIYARPSGLEQLRSGEFSRQFEELKLSVEVGRREPEELGKTGFGEIGET
SEG
PRD      hhhccccceeeccccccccchhhhhccccccccchhhhhhhhhhhhhcccccccccccccccc

SEQ      KHTGNEMEKEYE
SEG      .....
PRD      cccccccccccc

```

Prosite for DKFZphtes3\_15h1.3

PS000001	128->132	ASN_GLYCOSYLATION	PDCC000001
PS000001	438->442	ASN_GLYCOSYLATION	PDCC000001
PS000004	265->269	CAMP_PHOSPHO_SITE	PDCC000004
PS000004	605->609	CAMP_PHOSPHO_SITE	PDCC000004
PS000004	613->617	CAMP_PHOSPHO_SITE	PDCC000004
PS000004	636->640	CAMP_PHOSPHO_SITE	PDCC000004
PS000005	8->11	PKC_PHOSPHO_SITE	PDCC000005
PS000005	66->69	PKC_PHOSPHO_SITE	PDCC000005
PS000005	136->139	PKC_PHOSPHO_SITE	PDCC000005
PS000005	180->183	PKC_PHOSPHO_SITE	PDCC000005
PS000005	183->186	PKC_PHOSPHO_SITE	PDCC000005
PS000005	186->189	PKC_PHOSPHO_SITE	PDCC000005
PS000005	214->217	PKC_PHOSPHO_SITE	PDCC000005
PS000005	342->345	PKC_PHOSPHO_SITE	PDCC000005
PS000005	564->567	PKC_PHOSPHO_SITE	PDCC000005
PS000005	596->599	PKC_PHOSPHO_SITE	PDCC000005
PS000005	660->663	PKC_PHOSPHO_SITE	PDCC000005
PS000006	2->6	K2_PHOSPHO_SITE	PDCC000006
PS000006	66->70	K2_PHOSPHO_SITE	PDCC000006
PS000006	93->97	K2_PHOSPHO_SITE	PDCC000006
PS000006	171->175	K2_PHOSPHO_SITE	PDCC000006
PS000006	220->224	K2_PHOSPHO_SITE	PDCC000006
PS000006	277->281	K2_PHOSPHO_SITE	PDCC000006
PS000006	382->386	K2_PHOSPHO_SITE	PDCC000006
PS000006	392->396	K2_PHOSPHO_SITE	PDCC000006
PS000006	481->485	K2_PHOSPHO_SITE	PDCC000006
PS000006	507->511	K2_PHOSPHO_SITE	PDCC000006
PS000006	512->516	K2_PHOSPHO_SITE	PDCC000006
PS000006	542->546	K2_PHOSPHO_SITE	PDCC000006
PS000006	548->552	K2_PHOSPHO_SITE	PDCC000006
PS000006	628->632	K2_PHOSPHO_SITE	PDCC000006
PS000006	663->667	K2_PHOSPHO_SITE	PDCC000006
PS000007	506->515	TYR_PHOSPHO_SITE	PDCC000007
PS000008	119->125	MYRISTYL	PDCC000008
PS000008	132->138	MYRISTYL	PDCC000008
PS000008	213->219	MYRISTYL	PDCC000008

WO 01/12659

PCT/IB00/01496

PS00008	288->294	MYRISTYL	PDOC00008
PS00008	320->326	MYRISTYL	PDOC00008
PS00008	334->340	MYRISTYL	PDOC00008
PS00008	590->596	MYRISTYL	PDOC00008
PS00009	596->600	AMIDATION	PDOC00009
PS00009	603->607	AMIDATION	PDOC00009
PS00009	641->645	AMIDATION	PDOC00009

(No Pfam data available for DKFZphtes3\_15h1.3)

DKFZphtes3\_1515

group: cell structure and motility

DKFZphtes3\_1515 encodes a novel 717 amino acid protein with similarity to radial spokehead proteins.

The novel protein is similar to the *Chlamydomonas reinhardtii* radial spokehead protein of flagella or axoneme and to the *Strongylocentrotus purpuratus* sea urchin spermatozoa protein p63. This protein is important for the maintenance of a planar form of sperm flagellar beating. In addition, the novel protein contains a transferrin signature 1 for iron-binding. The new protein seems to be a part of the human radial spoke heads in spermatozoa.

BLAST results: No predictive prosite, pfam or SCOP motif.

The new protein can find application in modulating the structure of the human spermatozoa radial spoke head and modulation of sperm motility in men.

strong similarity to "radial spokehead" proteins

complete cDNA, complete cds, 1 EST hit (from a testis library)  
"radial spokehead" part of flagella in *Chlamydomonas*, this protein seems to be part of the sperm motor or tail

Sequenced by GBF

Locus: unknown

Insert length: 2478 bp

Poly A stretch at pos. 2452, polyadenylation signal at pos. 2433

```
1 CACCCCTGGCC CGCTCCCCGC GCCCTCCACG GGTAAACGGCC CCCTCTCTCG
51 GTGCTCAGAA ACCGGCCGGTG TCGACAGGTG GCTCTCGCTT GGCCTCCTTG
101 TCTGCAAGCC TTTCTCCTAG AGATCTCTGC CTCCTGGGGA ACCATGGGAG
151 ACCTGCCGCC CTACCCCTGAG CGCCCTCCCC AGCAGCCTCC GGGCCGGAGG
201 ACTTCTCAGG CCTCCACAGG CGCGACAGT CGGACCAAG CTCAGGCCCT
251 GGCAGCGGAG CCGAGAGAGA GCGAGCAGAT ACCTCCAGAC GCCAGCGGAA
301 AGCCCTCTGG TTGGTCACAG AGGGGCGAGC TGTCCCAACA GGAGAAGTTG
351 CTGATGGCCCC AGGTCTTTCCA GCGTGAAGAA GCGCCGGCTG GTGCATGGA
401 GTACCCATCT GTGAACACGG GCTTTCCCTC AGAGTCCAG CCTCAGCCTT
451 ACTCTGATGA AAGCAGGATG CAGGTCCGCG AGCTCACCAC CAGCCTAATG
501 CTGCAAGCGG TCCAGCAGGG CCAACGCCCT CTCTCCAGC AACTGGACCC
551 CACCTTCCAG GAGCCCCCAG TCAACCCCTT GGGCCAGTTC AACTCTTACC
601 AGACAGACCA GTTCTCTGAA GGTGCCCAGC ACGGGCCTTA CATAAGGGAT
651 GACCCCTGCC TTCACTTCTT GCGCTCTGAG CTGGGCTTCC CACACTACAG
701 TGCCCCAGGT CCGAGCCCGG AGCCTCTGGA GCTGGCCGTG CAGAACGCCA
751 AGCCCTACCT GCTGACAGAC AGCATCAATT GCGACCTCAG CCTGTACGAG
801 CACTGGTAA ATCTGCTGAC CAAGATCCTG AACCAAGCGG CTGAGGACCC
851 CTCTCTGTG CTGAGCTCTC TGAACCGCAC CACCGAGTGG GAGTGGTTCC
901 ACCCCAGACT GGACACGCTG CGGACGACCC CCGAGATGCA GCCCACCTAC
951 AAGATGGCGG AGAACACGAA GCGCTCTTTC ACCCGGAGTG GAGGCGGCAC
1001 TGAAGGCCAA CAGGAGATGG AGGAGCAGGT GGGGGAGACA CCAAGTCCCA
1051 ACATCATGGA GACTGCCTTC TACTTCGAGC AGGCCGCGGT CGGCTGAGC
1101 TCGGACGAGA GCTTCCGCAAT TTTCTGGGCC ATGAAACAGC TGSTGGAGCA
1151 GCAGCCCATC CACACCTCTC GCTTCTGGGG CAAGATCCTG GGAATCAAAC
1201 GCAGCTACCT GGTGCCCGAG GTGGAATTCC GGGAGGGCGA GGAGGAGGCA
1251 GAGGAGGAGG AGCTGCAGGA GATGACGGAA GGTGGCGAGG TCATGGAGGC
1301 GCACGGCGAG GAGGAGGGCG AGGAGGACGA GGAGAAGGCC GTGGACATCG
1351 TCCCTAAGTC CGTATGGAAG CCGCCGCCCC TGATCCCAAA GGAGGAGAGC
1401 CGCTCAGGCG CCAACAACTA CCTGTACTTT GTGTGCAAGC AGCCGGGCCT
1451 GCCATGGAGC CGGCTCCCCC ACCTCCTACC AGCCAGATGC GTGAACGCCC
1501 GAAAGATCAA GAAGTTCTTC ACAGGCTACC TGGACACGCC AGTCGTGAG
1551 TACCCACCCCT TCCCGGCGCA CGAGGCCAAC TACCTCGCGG CCCAGATAGC
1601 CCGCATCTCG GCCGCCACGC AGGTACAGCC GCTGGGCTTC TACCAGTTTA
1651 GTGAGGAGGA GGGCGACGAG GAGGAGGAAG GTGGTGTCTG GCGGACTCC
1701 TACGAGGAGA ACCCGGACTT CAGAGGSCATC CCCGTGCTGG AGCTGGTCGA
1751 CTCCATGGCC AACTGGGTGC ATCACACACA GCACATCCTG CCGCAGGGCC
1801 GCTGCACATT GGTGAACCCCT TTGCAGAGAA CAGAGGAGGA GGAGGACCTG
1851 GGGGAGGAGG AAGAGAAGGC AGATGAGGGG CAGAGATGCA GAAATCATGC
1901 GGTGGGCCCC CCACTGCTAA CGCCACTTTC AGAGACTGCA GAAATCATGC
1951 ACCTGGCACC CTGGACCAAC CGCCTGTCTT CAGCCTCTG CCGCAGTATC
2001 TCAGTGGCGG TTGTGCGCTC CAACCTCTGG CCGGGGGGCT ATGCTATGCG
2051 CAGTGGCAAA AAGTTTGAGA ACATCTACAT CGGCTGGGGT CACAAGTACA
2101 GCCCGGAGAG CTTCACCCCG GCGCTGCCAG CCCCATTCA ACAAGAGTAC
2151 CCGAGTGGCC CAGAGATCAT GGAGAACAGC CCGGAGGCC ACAGAGGAGG
2201 GCAGGCTCTG AAAGCAGCCC GAGGAAGAGG AGGAGACAGA TGACTGAGGC
2251 AGGAGGAGG CGAGGAGGAG GAGGAGGCGG AGGAGACAGA TGACTGAGGC
```

2301 CCACCCCTCTA GCCACTTTCC CCAAGCAGGT AGATAGCAAA TTTCCTCTTA  
2351 GAGGTAGTTA GCATGGATTA TATTTTCACT ATGTGCTTCC TGTCCTCCAGA  
2401 GGCAGGGAT AGAAAAGGAA GGCAACTGCT TCAAAATAAA TTCCTCCACG  
2451 GCATTAAAAA AAAAAAAGG AAAAAAAG

## BLAST Results

No BLAST result

## Medline entries

86251010:  
Molecular cloning and expression of flagellar radial spoke and dynein  
genes of  
Chlamydomona

81142496:  
Radial spokes of Chlamydomonas flagella: polypeptide composition and  
phosphorylation of  
stalk components.

9450971:  
Molecular cloning and characterization of a radial spoke head protein of sea urchin sperm  
axonemes: involvement of the protein in the regulation of sperm motility.

## Peptide information for frame 3

ORF from 144 bp to 2294 bp; peptide length: 717  
Category: strong similarity to known protein

1 MGDLPYPYPER PAQPPGRRRT SQASQRRHSR DQAQALAADP EERQIIPDA  
51 QRNAPGWSQR GSLSQQENLL MPQVFQAEAA RLGGMETPSV NTGFFSEFQP  
101 QPYSDESRMQ VAELTTSML QRLQQGSSSL FQQLDPTFQE PPVNPGLQFN  
151 LYQTDQFSEG AQHGPIYRDD PALQFLPSEL GFPHYSAQVP EPEPLELAVQ  
201 NAKAYLLQTS INCDLSLYEH LVNLLTKILN QRPEDPLSVL ESLNRTQWE  
251 WFHPKLDLTL DDPQMPTYK MAEKQKALFT RSGGGTEGEQ EMEEVGETP  
301 VPNIMETAFY FEQAGVGLSS DESFRIFLAM KOLVEQQPIH TCRFWGKILG  
351 IKRSYLVAEV EFREGEEAE EEEVEEMTEG GEVMEAHGEE EGEEDEEKAV  
401 DIVPKSVWKP PPVIPKEESR SGANKYLYFV CNEPGLPWTR LPHVTPAQIV  
451 NARKIKKFFT GYLDTPVVS Y PFFPGNEANY LRAQIARISA ATQVSPGLFY  
501 QFSEEEGDEE EGGAGRDSY EENPDFEGIP VLELVDSMAN WVHHTQHILP  
551 QGRCTWVWNL QKTEEEEDLG EEEKADEGP EEEVEQVGP LTPLESDAE  
601 IMHLAPWTTT LSCSLCPQYS VAVVRSNLWP GAYAYASGKK FENIYIGWGH  
651 KYSPESFNPA LPAPIQEQYYP SGPEIMEMSD PTVEEQALK AAQEALGALT  
701 EEEEEGEEEE EGEETDD

## BLASTP hits

Entry U73123.1 from database TREMBL:  
product: "radial spokehead"; Strongylocentrotus purpuratus radial  
spokehead mRNA, complete cds.  
Score = 1604, P = 7.4e-165, identities = 303/523, positives = 395/523

Entry B44498 from database PIR:  
radial spoke protein 6 - Chlamydomonas reinhardtii  
Score = 386, P = 3.4e-45, identities = 105/264, positives = 138/264

Alert BLASTP hits for DKFZphtes3\_15i5, frame 3

No Alert BLASTP hits found

Pedant information for DKFZphtes3\_15i5, frame 3

## Report for DKFZphtes3\_15i5.3

[LENGTH] 717  
[MW] 80913.61  
[pI] 4.36

[HOMOL] TREMBL:U73123\_1 product: "radial spokehead"; Strongylocentrotus purpuratus  
radial spokehead mRNA, complete cds. 1e-130  
[PROSITE] TRANSFERRIN\_1 1  
[PROSITE] MYRISTYL 5  
[PROSITE] AMIDATION 2  
[PROSITE] CAMP\_PHOSPHO\_SITE 2  
[PROSITE] CK2\_PHOSPHO\_SITE 14  
[PROSITE] TYR\_PHOSPHO\_SITE 1  
[PROSITE] GLYCOSAMINOGLYCAN 1  
[PROSITE] PKC\_PHOSPHO\_SITE 8  
[PROSITE] ASN\_GLYCOSYLATION 1  
[KW] All\_Alpha  
[KW] LOW\_COMPLEXITY 21.48 %

```

SEQ  MGDLPYPYPERPAQQPPGRRTSQASQRRHSRDQAQALAADPEERQQIPPDARNA PGWSQR
SEG  ....XXXXXXXXXXXXX.....
PRD  cccccccccccccccccccccchhhhhhhhhhhhhhhhhcccccccccccccccccccc

SEQ  GSLSQQENLLMPQVQAEARLGGMEYPSVNTGFPSEFQPPQPSYDESRMQVAELTTSML
SEG  .....XXXXX
PRD  cccchhhhhhhhhhhhhhhhhccccccccccccccccccccccccchhhhhhhhhhhhhhh

SEQ  QRLQOGSSLFQQLDPTFQEPVPLGQFNLYQTDQFSEGAHQGPYIRDDPALQFLPSEL
SEG  xxxxxxxxxxxxxxxx.....
PRD  hhhhhcccccccccccccccccccccccccccccccccccccccccccccccccccc

SEQ  GFPHYSAQVPEPELEAVQNAKAYLLQTSINCDLSLYEHLVNLTKILNQRPEPLSVL
SEG  .....
PRD  cccccccccccccchhhhhhhhhhhhhhhccccccccchhhhhhhhhhhhhhhccccchhh

SEQ  ESLNRTTQWEFHPKLDLTDPEMQPTYKMAEKQKALFTRSGGGTEGEQEMEEVEGETP
SEG  .....XXXXXXXXXXXXXXXXXXXXX...
PRD  hhhchhhhhccccccccccccccccccccchhhhhhhhhhhhhhhccccccccchhhhhhhcccc

SEQ  VPNIMETAFYFEQAGVGLSSDESFRILAMKQLVEQQPIHTCRFWGKILGIRSYLVAEV
SEG  .....XXX
PRD  cccchhhhhhhhhccccccccchhhhhhhhhhhhhhhccccchhhhhhhccccchhhhhhh

SEQ  EFREGEEAEVEEEMTEGGEVMEAHGEEEGEEDKEKAVDIVPKSVWKPVPVPIKEESR
SEG  xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx.....
PRD  hhhhhhhhhhhhhhhhhccccccccccccccccchhhheeecccccccccccccccccccc

SEQ  SGANKLYFVCNEPGLPWTRLPHVTPAQIVNARKIKKFTGYLTPVVSYPFPFNEANY
SEG  .....
PRD  cccceeeeeccccccccccccccccchhhhhhhhhhhhhhhccccccccccccccccchhh

SEQ  LRAQIARISAAQVSPGLGFYQFSEEGDEEEEGGAGRDSYEENPDFEGIPVLELVDSMAN
SEG  .....XXXXXXXXXXXXXXXXX.....
PRD  hhhhhhhhhhhhhccccccccceeeccccccccccccccccccccccccceeeccchhh

SEQ  WVHTQHILPQGRCTWVNPLOKTEEEEDLGESEKADGPEVEQEVGPPLTPLSEDAE
SEG  .....XXXXXXXXXXXXXXXXXXXXXXXXXXXXX.....
PRD  hhccccccccceeechhhhhhhhhccccchhhhhcccccccccccccccccccccccc

SEQ  IMHLAPWTRLSCSLCPQYSVAVVRSNLWPGAYAYASGKKFENIYIGWGHKYSPEFNP
SEG  .....
PRD  cccccccccccccccccccccceeeccccceeeccccccccceeecccccccccccccccc

SEQ  LPAPIQEQYPSGPEIMEMSDPTVEEEQALKAQEQALGATEEEEEEEEEEGEETDD
SEG  .....XXXXXXXXXXXXXXXXX...XXXXXXXXXXXXXXXXXXXXX...
PRD  cccccccccccccceeeccccchhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhcccc

```

## Prosites for DKF2phtes3\_1515.3

PS00001	244->248	ASN_GLYCOSYLATION	PDOC00001
PS00002	282->286	GLYCOSAMINOGLYCAN	PDOC00002
PS00004	18->22	CAMP_PHOSPHO_SITE	PDOC00004
PS00004	26->30	CAMP_PHOSPHO_SITE	PDOC00004
PS00005	24->27	PKC_PHOSPHO_SITE	PDOC00005
PS00005	58->61	PKC_PHOSPHO_SITE	PDOC00005
PS00005	258->261	PKC_PHOSPHO_SITE	PDOC00005
PS00005	268->271	PKC_PHOSPHO_SITE	PDOC00005
PS00005	323->326	PKC_PHOSPHO_SITE	PDOC00005
PS00005	341->344	PKC_PHOSPHO_SITE	PDOC00005
PS00005	608->611	PKC_PHOSPHO_SITE	PDOC00005
PS00005	637->640	PKC_PHOSPHO_SITE	PDOC00005
PS00006	64->68	CK2_PHOSPHO_SITE	PDOC00006
PS00006	137->141	CK2_PHOSPHO_SITE	PDOC00006

WO 01/12659

PCT/IB00/01496

PS00006	216->220	CK2_PHOSPHO_SITE	PDOC00006
PS00006	238->242	CK2_PHOSPHO_SITE	PDOC00006
PS00006	247->251	CK2_PHOSPHO_SITE	PDOC00006
PS00006	258->262	CK2_PHOSPHO_SITE	PDOC00006
PS00006	286->290	CK2_PHOSPHO_SITE	PDOC00006
PS00006	319->323	CK2_PHOSPHO_SITE	PDOC00006
PS00006	503->507	CK2_PHOSPHO_SITE	PDOC00006
PS00006	519->523	CK2_PHOSPHO_SITE	PDOC00006
PS00006	563->567	CK2_PHOSPHO_SITE	PDOC00006
PS00006	671->675	CK2_PHOSPHO_SITE	PDOC00006
PS00006	682->686	CK2_PHOSPHO_SITE	PDOC00006
PS00006	700->704	CK2_PHOSPHO_SITE	PDOC00006
PS00007	639->646	TYR_PHOSPHO_SITE	PDOC00007
PS00008	284->290	MYRISTYL	PDOC00008
PS00008	315->321	MYRISTYL	PDOC00008
PS00008	350->356	MYRISTYL	PDOC00008
PS00008	435->441	MYRISTYL	PDOC00008
PS00008	475->481	MYRISTYL	PDOC00008
PS00009	16->20	AMIDATION	PDOC00009
PS00009	637->641	AMIDATION	PDOC00009
PS00205	619->628	TRANSFERRIN_1	PDOC00182

(No Pfam data available for DKF2phtes3\_15i5.3)

DKF2phtes3\_15j18

group: testes derived

DKF2phtes3\_15j18 encodes a novel 148 amino acid protein without similarity to known proteins.

No informative BLAST results; No predictive prosite, pfam or SCOP motife.

The new protein can find application in studying the expression profile of testis-specific genes.

unknown

complete cDNA, complete cds, few EST hits

Sequenced by GBF

Locus: unknown

Insert length: 905 bp

Poly A stretch at pos. 839, polyadenylation signal at pos. 815

```
1 GTGATTCATA TGCTTCCATA GCAGGTGTCT GCTTCTGAGC CAAGCTCCCA
51 GGGCAGCGGA GCAGGCACCA ACCAGCATCC CAGGGGAGGG CACAGCTTGT
101 CCAGCTGGGA TGTTTGGGTG CCTGTGAGA TGCCCCAAGC CACCAACCCA
151 GCTTATCTCA GGAGAAGCCT CGGCGGCCCG TCTGCCGCCG TGGAGAGATG
201 TGCTACAGCA GCCGGGGGTG GGGGGAGAGG GTGGGCTTAG AATCTCTTGG
251 CAGGGAGCCC CCAAGAGCAG GGTGAGACCT GCCTTCATT CACCTGTCCC
301 CTTACAGTT CTGCAAAGCC AGCATTATCA TCCCTTTTCA GAAGGAGTGG
351 GCACTCAGGT GGAATGCCCT ACCCCAGTCC TGGGCTGGA AAGCGATATG
401 GCCAGGACTG CACCCACACC CTCATCCCTG CACCCCTTCC CTGCCCTGGGA
451 TTCTTCCAGC CCTGTGCACT GTGGAGCGCC TCTGCCTTCC GCTCATGGAG
501 GTTTCCCAAG GGCACGCGCT GAGGGCAGCT GGTCTCAGCC TGGGGCCGGG
551 TCCTAGTAAC TGCTCTCTT TGCTTTCCAG CCAGTGTITT GGGGTTTGAA
601 GTTGGAATCT TCAGCTACTG TCAAGAACAG CCACAAAAAT GTGTCACGAT
651 CAAGATCTTT GAGAGTCCAC CAATCAGGAG GCGTCTGTGA CAGTCGCTGT
701 CTTCTCAGAA CAGAATCCAC ACCCAGGATT CAACCCAAAT GATTTCTCAT
751 CAGGTGATTC TTGGTTGTAG CAAAGTTTCA GTGAATGTGG GTGAGTTTCT
801 GTTATGAATG TGGTCAATAA ATGTTATTG TGAAACTCTA AAAAAAAAAA
851 AAAAAAAAAA GGGCGCCGCT CTAGAGGATC CAAGCTTACG TACGCGAAAA
901 AAAAG
```

## BLAST Results

No BLAST result

## Medline entries

No Medline entry

## Peptide information for frame 2

ORF from 110 bp to 553 bp; peptide length: 148  
Category: putative protein

```
1 MFGCPVRCPK PPTQLISGEA SAARLPAPWRD VLQQPGVGGE GGLRISWQGA
51 PKSRVRPAFI SPVPFTVLQS QHYHFFSEGV GTQVECLTPV LRLESDMART
101 APHPSSLHPF PAWDSSSPVH CGAPLPSAHG GFPRARAEGS WSQPGAGS
```

## BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKF2phtes3\_15j18, frame 2

No Alert BLASTP hits found

Pedant information for DKF2phtes3\_15j18, frame 2

Report for DKF2phtes3\_15j18.2

```

(LENGTH)      148
(MW)           15665.78
(pI)           8.91
(PROSITE)      MYRISTYL      3
(CK2_PHOSPHO_SITE) 1
(KW)           Irregular

SEQ  MFGCPVRCPKPPTQLISGEASAARLPARVDVLPQPGVGGEGGLRISWQGAPKSRVRPAFI
PRD  ccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc

SEQ  SPVPFTVLQSQHYHPFSEGVGTQVECLTPVLRLESMDARTAPHPSLHPFPAWDSSSPVH
PRD  ccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc

SEQ  CGAPLPSAHGGFPARAEGSWSQPGAGS
PRD  ccccccccccccccccccccccccccccccccccccccc

```

Prosites for DKF2phtes3\_15j18.2

PS00006	82->86	CK2_PHOSPHO_SITE	PDOC00006
PS00008	38->44	MYRISTYL	PDOC00008
PS00008	42->48	MYRISTYL	PDOC00008
PS00008	49->55	MYRISTYL	PDOC00008

(No Pfam data available for DKF2phtes3\_15j18.2)



DKFZphtes3\_15j3  
-----

group: nucleic acid management

DKFZphtes3\_15j3 encodes a novel 743 amino acid protein with similarity to proteins with unknown function.

The novel protein contains a RNA recognition motif, predicted by Pfam and therefore binds to RNA. The protein is similar to YGR276c, a ribonuclease H of *S. cerevisiae*. Thus, the protein seems to be a new RNA-modifying protein.

The new protein can find application in modulating the RNA metabolism in human cells and as a tool for biotechnologic manipulations.

"44M2.3"; product, differences to genmodel, similarity to ribonuclease H

complete cDNA, complete cds, EST hits  
YGR276c = ribonuclease H  
differences to genmodel of 44M2.3

Sequenced by GBF

Locus: /map="16p11.2"

Insert length: 2695 bp  
Poly A stretch at pos. 2601, polyadenylation signal at pos. 2579

```
1 GCGGTTGTTG TTGGCAGCTG TGGCTAAGGA GGGGAGAACC TCTGCTCCCC
51 GCCCGTCTTC TCTTCTGCGT TTCCCGGGCT AGGGGGCGTG GGGAGTGGTT
101 TTAGGCGGCG AAGCCGCTCG GCAGCACCTT CCTTCTTTC CAGGCAGACG
151 CCCGTTGTAG CCGTTGGGGA ACCGTTGAGA ATCCGCCATG GAGCCAGAGA
201 GGAAGAGGAC CGAGAGACAC CCCAGGAAGG TCAGGGAAGG CAGGCAGGCC
251 CCAAAATAGC TGGTCGGGGC AGCTGAGGCG ATGAAAGCCG GTTGGGATCT
301 CGAGGAGAGT CAGCCCGAGG CCAAGAAAGC CCGCTTATCT ACCATTTTAT
351 TTAGTGACAA CTGTGAAGTA ACCCATGACC AGCTGTGTGA ATTGCTGAAG
401 TATGCAGTTC TGGGCAATC CAATGTTCCA AAACCCAGCT GGTGCCAGCT
451 TTTTCATCAA AACCACTAA ACAACGTAGT GGTTTTGTG CTGCAGGGAA
501 TGAGTCAGCT ACACTTTAC AGGTTCTATT TGGAGTTGG ATGCTTTCGA
551 AAAGCATTCG GACATAAAT CCGCTTGCTC CCACCATCAT CTGATTTTCT
601 AGCTGATGTT GTTGGGCTAC AAACCTGAACA AAGAGCTGGA GATCTGCCCA
651 AGACAATGGA AGGGCCCTTA CCTTCTAATG CAAAGCCCGC CATCAACCTT
701 CAGGATGATC CCATCATTCA AAGATATGGC TCTAAGAAAG TGGGCTTGAC
751 CAGATGCCCT CTGACAAAGG AGGAAATGAG AACGTTTAC TTTCATTAC
801 AAGGTTTTCC TGATTGTGAA AACTTTTAC TTACCAAAAT TAATGGTTCT
851 ATAGCAGACA ATAGTCTCTC CTTTGGACTT GACTGTGAAA GTGCGCTCAC
901 ATCCAAGGGG AGAGAGCTAA CACGCATCTC ACTGTTGCTT GAAGGAGGCT
951 GCTGTGTTAT GGATGAACCTG GTCAAACCTG AAAACAAGAT TCTGGACTAC
1001 CTCACCAAGT TTTCCGGGAT CACGAAGAAG ATTCTTAACC CAGTGACGAC
1051 CAAACTCAAA GATGTACAGA GGCAGTTAAA AGCACTGCTT CCTCTGATG
1101 CTGTGTTAGT GGGCCACTCC TTAGATTGGG ATCTCAGAGC ACTGAAAATG
1151 ATACATCCAT ATGTTATTGA TACATCGTTG CTTTATGTCA GAGAGCAGGG
1201 CAGAAGATTT AAGCTCAAAT TCTTAGCCAA AGTTATTTTG GGAAGGATA
1251 TACAGTGTCG AGACAGACTT GGTGATGATG CCACAGAAGA TGCTAGAACA
1301 ATCCTTGAAT TGGCTCGGTA TTCTCTAAG CATGGCCCAA AAAAGATTGC
1351 AGAATCAAA CTAGAGACAC TAGCTAATCA CCAAGAAATA CAAGCAGCAG
1401 GCCAAGAGCC TAAAAACACA GCAGAAGTAC TTCAGCACCC AAACACAAGT
1451 GTTTTAGAAT GCTTGGATTG AGTGGGTCAG AAGCTTCTTT TTTTGACCCG
1501 GGAGACAGAT GCTGGTGAAC TTCCATCTTC CAGAAATTGT CAAACTATTA
1551 AGTGTCTTTC AAATAAGAGG GTTCTTGAGC AGGCCAGAGT GGAATCCCC
1601 CTGTTTCCCT TCAGCATTGT TCAGTTCTCT TTTAAGGCTT TTTACCTGT
1651 CCTCACTGAG GAGATGAACA AAAGATGAG GATCAAGTGG ACAGAGATAT
1701 CAACGTGCTA TGCTGGGCCA TTTAGCAAAA ATTGCAATCT CAGGGCTCTG
1751 AAGAGGCTGT TTTAAAGCTT TGGCCAGTCT CAGTCAATGA CTTTGTCTCT
1801 TGAAACCCGT CAGGTGCAGA GGCCTGTGAC AGAGCTCACG CTTGATTGTG
1851 ACACCCCTGT GAATGAGCTG GAAGGAGATT CTGAAACCA AGGCTCTATA
1901 TATCTGTCTG GAGTGAGTGA AACCTTCAA GAACAGCTAT TGCAGGAGCC
1951 CCGCCTCTT CTTGGCCTGG AAGCTGTGAT CTTGCCTAAA GATCTTAAAA
2001 GTGGAAGACA GAAAAATAC TGTTTCTGTA AATTCAAAAG TTTTGGCAGT
2051 GCCCAGCAGG CCTCAACAT TCTCAGAGC AAGGACTGGA AGCTGAAAGG
2101 CAGGCATGCC CTAACCCCA GGCACCTCCA TGCTGGGCTC AGAGGCTTAC
2151 CACCTGAATC AACAGGCTC CCAGGGCTTC GTGTTGTACC TCCCCCTTT
2201 GAACAGGAGG CTTGCAGAC TCTGAACTG GACCACCCGA AGATAGCAGC
2251 CTGGCGCTGG AGCCGAAGA TTGAAAGCT CTACAACAGC TTGTGCCCGG
2301 GCACTCTCTG CCTCATCTG CTGCCAGGAA CCAAGAGCAC TCATGGTTCA
2351 CTCTCTGGTC TAGGACTGAT GGGAAATAAA GAGGAAGAAG AAAGCGCTGG
2401 CCCAGGCCCTG TGTTCTGTAG TCGGCTGACC ATGTTTCCAT GTGCCATTTC
```

```

2451 TTACCCCTTG TAGGCAATGG CAAAGAATGT GGTGAGGCTG TAGCCTCCCC
2501 AACCAGCAGA CAGTTTTATG GAAACTTGGT ATAGCAGCTA AAAGAGTTTA
2551 GTTTGTTTAT ATGGCATGTA TAAGTTTCA ATAAATGCCT AAAGTTCAAG
2601 CATAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA
2651 AGGGCGCCCG CTCTAAAGGA TCCAAGCTTA CGTACGCGAA AAAAG

```

## BLAST Results

No BLAST result

## Medline entries

No Medline entry

## Peptide information for frame 2

ORF from 188 bp to 2416 bp; peptide length: 743  
 Category: similarity to known protein

```

1 MEPEEREGTER HPRKVRERQ APNKLVGAAE AMKAGWDL EE SQPEAKKARL
51 STILFTDNCE VTHDQCELL KYAVLGKSNV PKPSWCQLFH QNHLNNVVVF
101 VLQGMSQLHF YRFYLEFGCL RKAFRHKFRL PPPSSDFLAD VVGLQTEQRA
151 GDLPKTMEGP LPSNAKAAIN LQDDPIIQKY GSKKVGLTRC LLTKEEMRTF
201 HFPLQGFPDC ENFLLTKCNG SIADNSPLFG LDCMCLTSK GRELTRISLV
251 AEGGCCVMEDE LVPENKILD YLTSFSGITK KILNPVTTKL KDVRQLKAL
301 LPPDAVLVGH SLDLDRALK MIHPYVIDTS LLYVREQRRR FKLFLAKVI
351 LGKDIQCPDR LGHATEDAR TLELARYFL KHGPKKIAEL NLEALANHQE
401 IQAAGQEPKN TAEVLQHPNT SVLECLDSVG QKLLFLTRET DAGELPSSRN
451 CQTIKCLSNK EVLEQARVEI PLFPFSIVQF SFKAFSPVLT EEMNKRMRK
501 WTEISTVYAG PFSKNCNLRA LKRLFKSFGP VQSMTFVLET RQVQRPVTEL
551 TLDCDTLVNE LEGDSENQGS IYLSGVSETF KEQLLQEPRL FLGLEAVILP
601 KDLKSGKQKK YCFLKFKSFG SAQALNILT GKDWKLGKRR ALTPRHLHAW
651 LRGLPPESTR LPGLRVVPPP FEQALQTLK LDHPKIAAWR WSRKIGKLYN
701 SLCPTGLCLI LLPGTKSTHG SLGGLGLMGI KEEESAGPG LCS

```

## BLASTP hits

No BLASTP hits available

## Alert BLASTP hits for DKF2phtes3\_15j3, frame 2

TREMBL:AC004381\_4 gene: "44M2.3"; product: "Unknown gene product";  
 Homo sapiens Chromosome 16 BAC clone CIT987SK-44M2, complete sequence.,  
 N = 2, Score = 1827, P = 2.1e-284

TREMBL:AF016430\_4 gene: "C05C8.5"; Caenorhabditis elegans cosmid  
 C05C8., N = 2, Score = 370, P = 1.7e-34

PIR:S64609 hypothetical protein YGR276c - yeast (Saccharomyces  
 cerevisiae), N = 2, Score = 334, P = 1.8e-27

TREMBLNEW:SPAC637\_9 gene: "SPAC637.09"; product: "putative  
 exonuclease"; S.pombe chromosome I cosmid c637., N = 3, Score = 326, P  
 = 2.8e-27

>TREMBL:AC004381\_4 gene: "44M2.3"; product: "Unknown gene product"; Homo  
 sapiens Chromosome 16 BAC clone CIT987SK-44M2, complete sequence.  
 Length = 547

## HSPs:

Score = 1827 (274.1 bits), Expect = 2.1e-284, Sum P(2) = 2.1e-284  
 Identities = 358/373 (95%), Positives = 358/373 (95%)

```

Query: 105 MSQLHFYRFYLEFGCLRKAFRHKFRLPPPSSDFLADVGLQTEQRAGDLPKTMEGPLPSN 164
      MSQLHFYRFYLEFGCLRKAFRHKFRLPPPSSDFLADVGLQTEQRAGDLPKTMEGPLPSN
Sbjct: 1 MSQLHFYRFYLEFGCLRKAFRHKFRLPPPSSDFLADVGLQTEQRAGDLPKTMEGPLPSN 60

Query: 165 AKAAINLQDDPIIQKYGSKKVGLTRCLLTKEEMRTFHFPPLQGFPDCENFLLTKCNGSIAD 224
      AKAAINLQDDPIIQKYGSKKVGLTRCLLTKEEMRTFHFPPLQGFPDCENFLLTKCNGSIAD
Sbjct: 61 AKAAINLQDDPIIQKYGSKKVGLTRCLLTKEEMRTFHFPPLQGFPDCENFLLTKCNGSIAD 120

```

Query:	225	NSPLFGLDCEM-----CLTSKGRELTRISLVAEGGCCVMDELVKPENKIL	269
		NSPLFGLDCEM                  CLTSKGRELTRISLVAEGGCCVMDELVKPENKIL	
Sbjct:	121	NSPLFGLDCEMARTTFNFISIGLVQAECLTSKGRELTRISLVAEGGCCVMDELVKPENKIL	180
Query:	270	DYLTFSFGITKKILNPVTTKLKDVRQLKALLPPDAVLVGHSLDLDLRALKMIHPYVIDT	329
		DYLTFSFGITKKILNPVTTKLKDVRQLKALLPPDAVLVGHSLDLDLRALKMIHPYVIDT	
Sbjct:	181	DYLTFSFGITKKILNPVTTKLKDVRQLKALLPPDAVLVGHSLDLDLRALKMIHPYVIDT	240
Query:	330	SLLYVREGRGRRFLKLFKAVLGKDIQCPRDLGHODATEDARTILELARYFLKHGPKKIAE	389
		SLLYVREGRGRRFLKLFKAVLGKDIQCPRDLGHODATEDARTILELARYFLKHGPKKIAE	
Sbjct:	241	SLLYVREGRGRRFLKLFKAVLGKDIQCPRDLGHODATEDARTILELARYFLKHGPKKIAE	300
Query:	390	LNLEALANHQEIQAAQGEPKNTAEVLQHFNTPSVLECLDSVGQKLLFLTRETDAAGLPSSR	449
		LNLEALANHQEIQAAQGEPKNTAEVLQHFNTPSVLECLDSVGQKLLFLTRETDAAGLPSSR	
Sbjct:	301	LNLEALANHQEIQAAQGEPKNTAEVLQHFNTPSVLECLDSVGQKLLFLTRETDAAGLPSSR	360
Query:	450	NCQTIKCLSNKEV 462	
		NCQTIKCLSNKEV	
Sbjct:	361	NCQTIKCLSNKEV 373	
Score = 929 (139.4 bits), Expect = 2.1e-284, Sum P(2) = 2.1e-284			
Identities = 175/179 (97%), Positives = 177/179 (98%)			
Query:	538	LETQRQRPVTETLTDCTLVNELEGDSENGSGIYLSGVSETFEKQLLOEQPRLFLGLEAV	597
		L++VQRPVTELTDDCTLVNELEGDSENGSGIYLSGVSETFEKQLLOEQPRLFLGLEAV	
Sbjct:	628	LSNKEVRQPVTETLTDCTLVNELEGDSENGSGIYLSGVSETFEKQLLOEQPRLFLGLEAV	427
Query:	598	ILPKDLKSGQKKYCFLKFKSFGSAQAALNLTGDKDWLKGRAHALTPRHHLAWRLGPLPE	657
		ILPKDLKSGQKKYCFLKFKSFGSAQAALNLTGDKDWLKGRAHALTPRHHLAWRLGPLPE	
Sbjct:	428	ILPKDLKSGQKKYCFLKFKSFGSAQAALNLTGDKDWLKGRAHALTPRHHLAWRLGPLPE	487
Query:	658	STRPLGLRVVPVFEEQALQTLKLDHPKIAAWRWSRKIGKLYNSLCPGTLCILLPGTK	716
		STRPLGLRVVPVFEEQALQTLKLDHPKIAAWRWSRKIGKLYNSLCPGTLCILLPGTK	
Sbjct:	488	STRPLGLRVVPVFEEQALQTLKLDHPKIAAWRWSRKIGKLYNSLCPGTLCILLPGTK	546

Pedant information for DKFZphtes3\_15j3, frame 2

Report for DKFZphtes3\_15j3.2

```

[LENGTH]      743
[MW]           83536.58
[pI]           8.87
[HOMOL]        TREMBL:AC004381_4 gene: "44M2.3"; product: "Unknown gene product"; Homo sapiens
Chromosome 16 BAC clone CIT9875K-44M2, complete sequence. 0.0
[FUNCAT]       01.03.16 polynucleotide degradation [S. cerevisiae, YGR276c] 4e-30
[FUNCAT]       99 unclassified proteins [S. cerevisiae, YLR107w] 3e-13
[FUNCAT]       05.04 translation (initiation, elongation and termination) [S. cerevisiae,
YGL094c] 1e-10
[FUNCAT]       04.05.05 mrna processing (5'-end, 3'-end processing and mrna degradation) [S.
cerevisiae, YGL094c] 1e-10
[FUNCAT]       03.22 cell cycle control and mitosis [S. cerevisiae, YOL080c] 2e-10
[PROSITE]      MYRISTYL 5
[PROSITE]      AMIDATION 1
[PROSITE]      CK2_PHOSPHO_SITE 8
[PROSITE]      TYR_PHOSPHO_SITE 1
[PROSITE]      GLYCOSAMINOGLYCAN 1
[PROSITE]      PKC_PHOSPHO_SITE 16
[PROSITE]      ASN_GLYCOSYLATION 2
[PFAM]         RNA recognition motif. (aka RRM, RBD, or RNP domain)
[KW]           Alpha Beta

```

[illegible]

SEQ	LPDDAVLVGHSLDLDRALKMIHPYVIDTSLLYVREGRRRFLKFLARVILGKDIOCPDR
PRD	cc
SEQ	LGHDAATECARTLILRLARYFLKHGPKKIAELNLEALANHOEQOAAQGEPKNTAEVLQHPNT
PRD	cc
SEQ	SVLECLDSVQKLLFLTRETDAEGLPSSSRNMCOTIKCLSNKEVLEQARVEIPLFFSVSIQVF
PRD	cc
SEQ	SFKAFSPVLTEEMNKRMIKWTETISTVYAGFSPKNCNLRALKRFLKSGFVQSMTFVLET
PRD	cc
SEQ	RQVRPVTETLDCDCTLVNELEGDSENGQISLYSGVSETFKEQLLQHPRLFLGLAEVAILP
PRD	cc
SEQ	KDLKSGKQKKYCYFLKFKSFGSAQQAALNLTGDKWLKGRHALTPRHLHAWLGLPPESTR
PRD	cc
SEQ	LPGLRRVPPPFEEQAELQTLKDHPKIAAWRWSRIGKLYNSCLPGTCLLILPGTKSTHG
PRD	cc
SEQ	SLSGCCGHHGKEEESAGPGLCS
PRD	cccccccccccccccccccccccc

Prosite for DKFZphtes3\_15j3.2

PS000001	219->223	ASN_GLYCOSYLATION	PDCC000001
PS000001	419->423	ASN_GLYCOSYLATION	PDCC000001
PS000002	723->727	GLYCOSAMINOGLYCAN	PDCC000002
PS000005	8->11	PKC_PHOSPHO_SITE	PDCC000005
PS000005	182->185	PKC_PHOSPHO_SITE	PDCC000005
PS000005	238->241	PKC_PHOSPHO_SITE	PDCC000005
PS000005	279->282	PKC_PHOSPHO_SITE	PDCC000005
PS000005	287->290	PKC_PHOSPHO_SITE	PDCC000005
PS000005	447->450	PKC_PHOSPHO_SITE	PDCC000005
PS000005	453->456	PKC_PHOSPHO_SITE	PDCC000005
PS000005	458->461	PKC_PHOSPHO_SITE	PDCC000005
PS000005	481->484	PKC_PHOSPHO_SITE	PDCC000005
PS000005	579->582	PKC_PHOSPHO_SITE	PDCC000005
PS000005	605->608	PKC_PHOSPHO_SITE	PDCC000005
PS000005	630->633	PKC_PHOSPHO_SITE	PDCC000005
PS000005	643->646	PKC_PHOSPHO_SITE	PDCC000005
PS000005	658->661	PKC_PHOSPHO_SITE	PDCC000005
PS000005	678->681	PKC_PHOSPHO_SITE	PDCC000005
PS000005	692->695	PKC_PHOSPHO_SITE	PDCC000005
PS000006	41->45	CK2_PHOSPHO_SITE	PDCC000006
PS000006	193->197	CK2_PHOSPHO_SITE	PDCC000006
PS000006	221->225	CK2_PHOSPHO_SITE	PDCC000006
PS000006	371->375	CK2_PHOSPHO_SITE	PDCC000006
PS000006	421->425	CK2_PHOSPHO_SITE	PDCC000006
PS000006	458->462	CK2_PHOSPHO_SITE	PDCC000006
PS000006	579->583	CK2_PHOSPHO_SITE	PDCC000006
PS000006	630->634	CK2_PHOSPHO_SITE	PDCC000006
PS000007	370->379	TYR_PHOSPHO_SITE	PDCC000007
PS000008	27->33	MYRISTYL	PDCC000008
PS000008	186->192	MYRISTYL	PDCC000008
PS000008	575->581	MYRISTYL	PDCC000008
PS000008	714->720	MYRISTYL	PDCC000008
PS000008	720->726	MYRISTYL	PDCC000008
PS000009	337->341	AMIDATION	PDCC000009

Pfam for DKFZphtes3 15j3.2

HMM_NAME	RNA recognition motif. (aka RRM, RBD, or RNP domain)	
HMM	*IYVGNLPWDtTEEDLrDlFsQFGpIvsIRmMrDRtGRSGrGFAfVEFED	
	IY+ +++ +T +E+L + + F + + + +D G+ + ++F + ++	
Query	571 IYLSGVs-ETfKEQLQLEPRfLGLAEVILPKDLKSGRKQKYcFLKfKS	618
HMM	EEDAekAIdemNG..meFmGrIRv*	
	+A+ A+ + G ++ GR +	
Query	619 FGSAQQALNILTgKDWKLKGRHALT	643

DKF2phtes3\_15k11

group: signal transduction

DKF2phtes3\_15k11 encodes a novel 958 amino acid protein C-terminal identical with human KIAA0781 protein and high similarity to protein kinases.

The novel protein contains a protein kinase ATP-binding region signature and a serine/threonine protein kinase active-site signature. The related murine kinase was cloned from the myocardium of the developing heart.

The new protein can find application in modulation of intracellular signal pathways dependent on this kinase.

KIAA0781, 5' extension

complete cDNA, complete cds, potential start at Bp 97, EST hits

Sequenced by GBF

Locus: /map="11"

Insert length: 4868 bp

Poly A stretch at pos. 4798, polyadenylation signal at pos. 4776

```
1 GAGCAAGCGG AGCGGCCGTC GCCCAAGCCA AGCGCGCGTG CCAACCCCTCC
51 CGCCCGCCCG CGCTCCTGTC CGCCGTGTCT AGCAGCGGGG CCCAGCATGG
101 TCATGGCGGA TGGCCCGAGG CACTTGCAGC GCGGGCCGGT CCGGGTGGGG
151 TTCTACGACA TCGAGGGCAC GCTGGGCAAG GGCACCTTCG CTGTGGTGAA
201 GCTGGGGCGG CACCGGATCA CCAAGACGGA GGTGGCAATA AAAATAATCG
251 ATAAGTCTCA GCTGGATGCA GTGAACCTTG AGAAAACTTA CCGAGAAGTA
301 CAAATAATGA AAATGTTAGA CCACCCCTAC ATAATCAAAC TTTATCAGGT
351 AATGGAGACC AAAAGTATGT TGTACCTTGT GACAGAATAT GCCAAAAATG
401 GAGAAATTTT TGACTATCTT GCTAATCATG GCCGGTTAAA TGAGTCTGAA
451 GCCAGGCGAA AATCTGGGCA AATCCTGTCT GCTGTTGATT ATTGTCATGG
501 TCGGAAGATT GTGCACCGTG ACCTCAAAGC TGAAAACTTC CTGCTGGATA
551 ACAACATGAA TATCAAAATA GCAGATTTCG GTTTTGGAAA TTTCTTTAAA
601 AGTGGTGAAC TGCTGGCAAC ATGGTGTGGC AGCCCCCTTT ATGCAGCCCC
651 AGAAGTCTTT GAAGGGCAGC AGTATGAAG ACCACAGCTG GACATCTGGA
701 GTATGGGAGT TGTCTTTTAT GTCCCTGTCT GTGGAGCTCT GCCCTTTGAT
751 GGACCGGACT TTCCAATTTT GAGGCGAGGG GTTCTGGAAG GAAGATTCCG
801 GATTCCGTAT TTCTATGTCA AAGATTGCGA GCACCTTATC CGAAGGATGT
851 TGGTCCTAGA CCCATCCAAA CGGCTAACCA TAGCCCAAT CAAGGAGCAT
901 AAATGGATGC TCATAGAAAT TCCTGTCCAG AGACCTGTTT TCTATCCACA
951 AGAGCAAGAA AATGAGCCAT CCATCGGGGA GTTTAATGAG CAGGTTCTGC
1001 GACTGATGCA CAGCCTTGGA ATAGATCAGC AGAAAACCAT TGAGTCTTTG
1051 CAGAACAAGA GCTATAACCA CTTTGTCTGC ATTTATTTCT TGTGGTGGA
1101 GCGCCTGAAA TCACATCGGA GCAGTTTCCC AGTGGAGCAG AGACTTGATG
1151 GCGGCCAGCG TCGGCCTAGC ACCATTGCTG AGCAACAGT TGCCAAGGCA
1201 CAGACTGTGG GGCTCCAGT GACCATGCAT TCACCGAACA TGAGGCTGCT
1251 GCGATCTGCC CTCTCCCCCT AGGCATCCAA CGTGGAGGCC TTTTCATTTC
1301 CAGCATCTGG CTGTGAGGCG GAAGCTGCAT TCATGGAAGA AGAGTGTGTG
1351 GACACTCCAA AGGTCAATGG CTGTCTGCTT GACCCTGTGC CTCCTGTCTT
1401 GGTGCGGAAG GGATGCCAGT CACTGCCCAG CAACATGATG GAGACCTCCA
1451 TTGACGAAGG GCTGGAGACA GAAGGAGAGG CCGAGGAAGA CCCCCTCAT
1501 GCCTTTGAGG CATTTCAGTC CACACGCAGC GGGCAGAGAC GGCACACTCT
1551 GTCAGAAGTG ACCAATCAAC TGGTCGTGAT GCCTGGGGCA GGGAAAAATT
1601 TCTCCATGAA TGACAGCCCC TCCCTTGACA GTGTGGACTC TGAGTATGAT
1651 ATGGGGTCTG TTCAGAGGGA CCTGAACTTT TTGGAAGACA ACCCTTCCCT
1701 TAAGGACATC ATGTTAGCCA ATCAGCCTTC ACCCCGATG ACATCTCCCT
1751 TCATAAGCCT GAGACCTACC AACCAGCCA TGCAGGCTCT GAGCTCCCG
1801 AAACGAGAGG TCCACAACAG GTCTCCAGTG AGCTTCAGAG AGGGCCGCG
1851 AGCATCAGAT ACCTCCCTCA CCCAGGGAAT TGTAGCATT AGACAACTC
1901 TTCAGAATCT GGCTAGAACC AAAGGAATTC TAGAGTTGAA CAAAGTGCAG
1951 TTGTTGTATG AACAAATAGG ACCGGAGGCA GACCCTAACC TGGCGCCGCG
2001 GGCTCCTCAG CTCAGGACC TTGCTAGCAG CTGCCCTCAG GAAGAAGTTT
2051 CTCAGCAGCA GGAAGCGTC TCCACTCTCC CTGCCCGCT GCATCCCCAG
2101 CTGTCCCCAC GGCAGAGCCT GGAGACCCAG TACCTGCAGC ACAGACTCCA
2151 GAAGCCGAGC CTTCTGTCAA AGGCCAGAA CACCTGTGCA CTTTATTGCA
2201 AAGAACCACC GCGGAGCCTT GAGCAGCAGC TGCAGGAACA TAGGCTCCAG
2251 CAGAAGCGAC TCTTTCTTCA GAAGCAGTCT CAACTGCAGG CCTATTTTAA
2301 TCAGATGCAG ATAGCAGAGA GCTCCTACCC ACAGCCAAGT CAGCAGCTGC
2351 CCCCCTCCCG CCAGGAGACT CCACCGCTT CTAGCAGGC CCCACCGTTC
2401 AGCCTGACCC AGCCCTGAG CCCCCTGCT GAGCCTTCT CCGAGCAGAT
2451 GCAATACAGC CTTTCTCTCA GCCAGTACCA AGAGATGCAG CTTAGCCCC
2501 TGCCCTCCAC TTCCGGTCCC CGGGCTGCTC CTCTCTGCC CACGACGCTA
2551 CAGCAGCAGC AGCCGCCACC GCCACCACC CCTCCACCAC CACGACAGCT
2601 AGGAGCTGCC CCAGCCCCCT TACAGTTCTC CTATCAGACT TGTAGCTGCT
```

```
2651 CAAGCGCTGC TTCCCTGCG CCAGACTATC CCACTCCCTG TCAGTATCCT
2701 GTGGATGGAG CCCAGCAGAG CGACCTAACG GGGCCAGACT GTCCCAAGAG
2751 CCCAGGACTG CAAGAGGGCC CCTCCAGCTA CGACCCACTA GCCCTCTCTG
2801 AGCTACCTGG ACTCTTTGAT TGTGAAATGC TAGACGCTGT GGATCCACAA
2851 CACAACGGGT ATGTCTTGGT GAATTAGTCT CAGCACAGGA ATTGAGGTGG
2901 GTCAGGTGAA GGAAGAGTGT ATGTCTCTAT TTTTATTCCA GCCTTTTAAA
2951 TTTAAAGCTT ATTTTCTTGC CCTCTCCCTA ACGGGGAGAA ATCGAGCCAC
3001 CCAACTGGAA TCAGAGGGTC TGGCTGGGGT GGATGTTGCT TCCCTCTGGT
3051 TCTGCCCCAC CACAAAGTTT TCTGTGGCAA GTGCTGGAAC ATAGTTGTAG
3101 GCTGAGGCTC CTGCCCTTCG GTCGAGTGGA GCAAGCTCTC GAGGCGAGCA
3151 CTGACAAATG TGTTCCTAAG AAGACATTCA GACCCAGGTC TTATGCAGGA
3201 TTACATCCGT TTATATCAAA GGGCAACCTT GGTGAAAGCA GAAAGGTTGT
3251 GTGCTATTGC ATATATATGG GGGAAAAGGC AATATATTTT TCAGTGAAGC
3301 TGAGCAACCA CATATTGCTA CAAGGCAAAAT CAAGAAGACA TCAGGAAATC
3351 AGATGCACAG GAAATAAAGG AAAGCTGTGC TTTGTCTTGG AATCCTAAGT
3401 TCTTAGCTGC TGATGCAAGT TGTCCCCCAA GGCCATCACA AAGCAGTGGG
3451 GCATGAGCTG TGTTCAGGG GCCACTAAAT AACAGCTGGT ACTGACCCCA
3501 GAAACCGCCT TCATCTCCAT TCGGAAGCAG GTGACACACC CCTTCAGAAG
3551 GTGCCCTGGG TTGCCGAGTG TCAGAAATATA CTCAGGACTC CAGAGGTGTC
3601 ACACGTGGAA CTGACAGGAG ACCCGCCACC GTGGAGGCGAG GGGGCAAGAA
3651 ACTCAAGAAC GCATCAAGAG CACCAGCCCT GGGCCAGGGA AGACAGGCTC
3701 TTCTGCGAGT TTCTCGTGGA CACTGTGGCG TTGGGGGCGAG TCGGTTCCCA
3751 GGGTACCTGT TGTCTCTTTT CCGATGTAAT AACTACTTTG ACCTTACACT
3801 ATATGTTGCT AGTAGTTTAT TGAGCTTTGT ATATTGGAG AGTTTCATAT
3851 AGGGCTTAGA GATTTTAAGG ACATGATAAA TGAACCTTTT TGTCCTATGT
3901 GAAGTGGTAG TGGGTGCTCT TTCCCCCAGA TCATGCTTTA ATTCTTTCTT
3951 TTCTGTAGAA ACCAACAGTT TCCATTATG TCAATGCTAA ATCCAAAGTC
4001 ACTTCAGAGT TTGTTTCCCA CCAATGTGGG ATCAGCATTC TTAATTTCTG
4051 TAAAGTTTGG ACTTGTAAAG AAATGTTCAA GTATTACAGC AATATTCAAA
4101 GAAAGAACCA CAGATGTGTT AACCATTAA GCAGATCATC TGCCAAACAT
4151 TATATTACTA ATAAACTTAA ACCAACACTT ACAATTCAGT CATCAAAGTA
4201 AGTAAAAATT AGATGCTACA GCTAGCTAAC TGTATCCCTA GAAATGATGA
4251 ATAATTGGCC ATTTGGACAG TTAACATCCA GGTGTTACAA AGTCAGTGT
4301 AATTCTAAAG ATGATCATTT CTGCCCTTTA GAATGGCTTG TCCCATCAGC
4351 AGATGAATGT GTTAAGCACA AAGCATCTTC CTTAAAGCAC AAAGAGAGGG
4401 ACTAACTGAT GCTGCATCTA GAAAACACCT TTAAGTTGCC TTTCTCTTT
4451 GTAGTTAGCG TTCAGGCAGG TGACGTGTGG AAAGTCTAGG GGGTTCCATT
4501 CTGGCCATGC GAGCCGAGCT CTTACCAACG TCGGTAACTT GAGCAGTCCC
4551 TGTGCTGGC CAGAGACTGC CTGGTCCGCA CGGCTCACCA TGGGTGCCAG
4601 GATGCTTCGC AGAGGCAGTG TGCTCACGGT TGGACTTGGT GTGAGTGGGA
4651 AAGGGCAGTG TGGGGACTGT CATTTTTGTG ATTTAATAAC ACACAGTGAA
4701 AATCCAGGAA GAATGAATTA AGCTTCTTCT GGGAGTTGTT TATTCCTGCT
4751 CGTGCTTAAG ATTGATGATT TCGTGAAATA AAGAACATCA TTTCAATTA
4801 AAAAAAAGG AAAAAAGG CGGCCGCTCT AGAGGATCCA AGCTTACGTA
4851 CGCGTAAAAA AAAAAAG
```

## BLAST Results

Entry HSG4921 from database EMBL:  
human STS SHGC-37164.  
Score = 1605, P = 1.9e-66, identities = 349/369

Entry AB018324 from database EMBL:  
Homo sapiens mRNA for KIAA0781 protein, partial cds.  
Score = 10725, P = 0.0e+00, identities = 2145/2145

## Medline entries

No Medline entry

## Peptide information for frame 1

ORF from the beginning to 2874 bp; peptide length: 959  
Category: known protein

```
1 EQAERPSPKP SRAANPPARP RSCPPCLAAG PSMVMADGPR HLQGPVVRVG
51 FYDIEGLTGG GNFAVVKLGR HRITKTEVAI KIIDKSQDLA VNLEKIYREV
101 QIMKMLDHPH IIKLYQVNET KSMILYLVTEY AKNGEIPDYL ANHGRNLNESE
151 ARRKFWQILS AVDYCHGRKI VHRDLKAENL LLDNNMNIKI ADFGFGNFFK
201 SGELLATWCG SPPYAAPPEVF EGQQYEGPQL DIWSMGVVLY VLVCGALPFD
251 GPTLPILRQR VLEGRFRIPY FMSDECHELI RRMVLVDFSK RLITIAQIKH
```

```

301 KWMLEIEVPVQ RPYLYPQEQE NEPSIGEFNE QVRLMHSLG IDQKTIESTL
351 QNKSYNHFAA IYFLVVERLK SHRSSFPVEQ RLDGRQRRPS TIAEQTVAKA
401 QTVGLPVTHM SPNMRLLRSA LLPQASNVEA FSFPASGCQA EAAFMEEECV
451 DTPKVNGLCL DPVFPVVRK GCQSLPSNM ETSIDEGLET EGEAEEDPAH
501 AFEAFQSTRS GQRRHTLSEV TNQLVVMFGA GKIFSMNDSP SLDSDSEYD
551 MGSVQDLNLF LEDNPSLKDI MLANQPSPRM TSPFISLRPT NPAMQALSSQ
601 KREVNHRSPV SFREGRRASD TSLTQGI VAF RQHLQNLART KGILELNKVPQ
651 LLYEQIGPEA DPNLAPAPQ LQDLASSCPQ EEVSQQQESV STLPASVHPQ
701 LSPRQSLQEQ YLQHLRQKPS LLSKAQNTCQ LYCKEPPRSL EQQLQEHRLQ
751 QKRLFLQKQS QLQAYFNMQ IAESSYPQPS QQLPLPRQET PPPSQAPPF
801 SLTQPLSPVL EPSSEQMYS PFLSQYQEQ LQPLPSTSGP RAAPPLPTQL
851 QQQQPPPPPP PPPPRQPGAA PAPLOFSYQT CELPSAASPA PDYPTPCQYP
901 VDAQQSDLT GPDCPRSPGL QEAPSSYDPL ALSELPLGLFD CEMLDAVDQ
951 HNGYVLVN

```

## BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphtes3\_15kl1, frame 1

No Alert BLASTP hits found

Pedant information for DKFZphtes3\_15kl1, frame 1

## Report for DKFZphtes3\_15kl1.1

```

[LENGTH]      926
[MW]           103915.77
[pI]           5.70
[HOMOL]        TREMBL:AB018324_1 gene: "KIAA0781"; product: "KIAA0781 protein"; Homo sapiens
mRNA for KIAA0781 protein, partial cds. 0.0
[FUNCAT]       01.05.04 regulation of carbohydrate utilization [S. cerevisiae, YDR477w]
8e-76
[FUNCAT]       11.01 stress response [S. cerevisiae, YDR477w] 8e-76
[FUNCAT]       30.03 organization of cytoplasm [S. cerevisiae, YDR477w] 8e-76
[FUNCAT]       98 classification not yet clear-cut [S. cerevisiae, YCL024w] 4e-58
[FUNCAT]       03.25 cytokinesis [S. cerevisiae, YDR507c] 3e-56
[FUNCAT]       03.04 budding, cell polarity and filament formation [S. cerevisiae, YDR507c]
3e-56
[FUNCAT]       30.02 organization of plasma membrane [S. cerevisiae, YDR122w] 1e-53
[FUNCAT]       03.22 cell cycle control and mitosis [S. cerevisiae, YKL101w] 3e-53
[FUNCAT]       30.10 nuclear organization [S. cerevisiae, YKL101w] 3e-53
[FUNCAT]       99 unclassified proteins [S. cerevisiae, YPL141c] 5e-51
[FUNCAT]       03.19 recombination and dna repair [S. cerevisiae, YPL153c] 3e-42
[FUNCAT]       03.22.01 cell cycle check point proteins [S. cerevisiae, YPL153c] 3e-42
[FUNCAT]       10.99 other signal-transduction activities [S. cerevisiae, YPL153c] 3e-42
[FUNCAT]       11.04 dna repair (direct repair, base excision repair and nucleotide excision
repair) [S. cerevisiae, YPL153c] 3e-42
[FUNCAT]       03.01 cell growth [S. cerevisiae, YFR014c] 5e-42
[FUNCAT]       03.16 dna synthesis and replication [S. cerevisiae, YMR001c] 2e-34
[FUNCAT]       03.10 sporulation and germination [S. cerevisiae, YGL180w] 1e-27
[FUNCAT]       08.13 vacuolar transport [S. cerevisiae, YGL180w] 1e-27
[FUNCAT]       06.13.04 lysosomal and vacuolar degradation [S. cerevisiae, YGL180w] 1e-27
[FUNCAT]       10.02.11 key kinases [S. cerevisiae, YBL105c] 3e-26
[FUNCAT]       04.99 other transcription activities [S. cerevisiae, YER129w] 3e-26
[FUNCAT]       02.19 metabolism of energy reserves (glycogen, trehalose) [S. cerevisiae,
YPL031c] 1e-23
[FUNCAT]       01.04.04 regulation of phosphate utilization [S. cerevisiae, YPL031c]
1e-23
[FUNCAT]       04.05.01.04 transcriptional control [S. cerevisiae, YPL031c] 1e-23
[FUNCAT]       03.13 meiosis [S. cerevisiae, YOR351c] 2e-23
[FUNCAT]       10.05.11 key kinases [S. cerevisiae, YHL007c] 8e-21
[FUNCAT]       03.07 pheromone response, mating-type determination, sex-specific proteins
[S. cerevisiae, YHL007c] 8e-21
[FUNCAT]       09.01 biogenesis of cell wall [S. cerevisiae, YPL140c] 2e-20
[FUNCAT]       10.03.11 key kinases [S. cerevisiae, YLR113w] 7e-20
[FUNCAT]       04.05.01.01 general transcription activities [S. cerevisiae, YDL108w]
3e-19
[FUNCAT]       10.05.09 regulation of g-protein activity [S. cerevisiae, YBL016w] 2e-18
[FUNCAT]       10.04.11 key kinases [S. cerevisiae, YLR362w] 3e-18
[FUNCAT]       04.03.99 other trna-transcription activities [S. cerevisiae, YOR061w]
4e-18
[FUNCAT]       06.07 protein modification (glycosylation, acylation, myristylation,
palmitoylation, farnesylation and processing) [S. cerevisiae, YFL033c] 4e-17
[FUNCAT]       05.07 translational control [S. cerevisiae, YDR283c] 2e-16
[FUNCAT]       01.02.04 regulation of nitrogen and sulphur utilization [S. cerevisiae,
YNL183c] 2e-14

```

{FUNCAT} 08.99 other intracellular-transport activities [S. cerevisiae, YNL183c]  
 2e-14  
 {FUNCAT} 09.04 biogenesis of cytoskeleton [S. cerevisiae, YNL020c] 5e-14  
 {FUNCAT} c energy conversion [M. genitalium, MG109] 2e-12  
 {FUNCAT} 30.09 organization of intracellular transport vesicles [S. cerevisiae,  
 YBR097w] 1e-10  
 {FUNCAT} 08.07 vesicular transport (golgi network, etc.) [S. cerevisiae, YBR097w]  
 1e-10  
 {FUNCAT} 30.08 organization of golgi [S. cerevisiae, YBR097w] 1e-10  
 {FUNCAT} 06.04 protein targeting, sorting and translocation [S. cerevisiae, YBR097w]  
 1e-10  
 {FUNCAT} 10.04.99 other nutritional-response activities [S. cerevisiae, YJR059w]  
 4e-09  
 {FUNCAT} 01.06.10 regulation of lipid, fatty-acid and sterol biosynthesis [S.  
 cerevisiae, YHR079c] 1e-07  
 {FUNCAT} 30.07 organization of endoplasmatic reticulum [S. cerevisiae, YHR079c]  
 1e-07  
 {FUNCAT} 08.19 cellular import [S. cerevisiae, YNL154c] 2e-04  
 {BLOCKS} BL00415A Synapsins proteins  
 {BLOCKS} BL00239B Receptor tyrosine kinase class II proteins  
 {BLOCKS} BL00107A Protein kinases ATP-binding region proteins  
 {SCOP} dfgl\_ 5.1.1.1.9 MAP kinase Erk2 [rat Rattus norvegicus 3e-78  
 {SCOP} dlwfc\_ 5.1.1.1.8 MAP kinase p38 [human (Homo sapiens) 1e-81  
 {SCOP} dikoa\_2 5.1.1.1.7 (1-350) Twitchin, kinase domain [Caenorhabditis 5e-89  
 {SCOP} dikoba\_ 5.1.1.1.6 Twitchin, kinase domain [california sea har 5e-86  
 {SCOP} diphk\_ 5.1.1.1.5 gamma-subunit of glycogen phosphorylase kinas 3e-80  
 {SCOP} dlirk\_ 5.1.1.2.4 insulin receptor [Human (Homo sapiens) 6e-70  
 {SCOP} diapme\_ 5.1.1.1.4 cAMP-dependent PK, catalytic subunit [mouse (Mu 1e-95  
 {SCOP} dlfgka\_ 5.1.1.2.3 Fibroblast growth factor receptor 1 [human (Hom 7e-71  
 {SCOP} dlydse\_ 5.1.1.1.3 cAMP-dependent PK, catalytic subunit [bovine (Bo 2e-96  
 {SCOP} dlfmk\_3 5.1.1.2.2 (168-437) c-src tyrosine kinase [human (Hom 2e-72  
 {SCOP} dlcdka\_ 5.1.1.1.2 cAMP-dependent PK, catalytic subunit [pig (Su 5e-97  
 {SCOP} d2hckb3 5.1.1.2.1 (167-437) Haemopoietic cell kinase Hck [huma 2e-68  
 {SCOP} dlcsn\_ 5.1.1.1.11 Casein kinase-1, CK1 [Schizosaccharomyces pombe 3e-53  
 {SCOP} dljsua\_ 5.1.1.1.1 Cyclin-dependent PK [Human (Homo sapiens) 3e-78  
 {SCOP} dickia\_ 5.1.1.1.10 Casein kinase-1, CK1 [rat (Rattus norvegicus) 1e-58  
 {EC} 2.7.1.117 Myosin-light-chain kinase 3e-49  
 {EC} 2.7.1.109 [Hydroxymethylglutaryl-CoA reductase(NADPH)] kinase 4e-78  
 {EC} 2.7.1.38 Phosphorylase kinase 3e-41  
 {EC} 2.7.1.37 Protein kinase 7e-45  
 {EC} 2.7.1.123 Ca2+/calmodulin-dependent protein kinase 5e-42  
 {EC} 2.7.1.128 [Acetyl-CoA carboxylase] kinase 4e-78  
 {PIRKEW} phosphotransferase 3e-93  
 {PIRKEW} nucleus 2e-74  
 {PIRKEW} calcium 2e-40  
 {PIRKEW} transferase 3e-33  
 {PIRKEW} duplication 2e-32  
 {PIRKEW} tandem repeat 7e-45  
 {PIRKEW} phorbol ester binding 4e-33  
 {PIRKEW} zinc 4e-33  
 {PIRKEW} ion transport 1e-32  
 {PIRKEW} cell cycle control 1e-45  
 {PIRKEW} serine/threonine-specific protein kinase 2e-97  
 {PIRKEW} oncogene 1e-34  
 {PIRKEW} phospholipid binding 2e-32  
 {PIRKEW} autophosphorylation 2e-74  
 {PIRKEW} brain 6e-36  
 {PIRKEW} heterotetramer 8e-38  
 {PIRKEW} mitosis 1e-45  
 {PIRKEW} polymer 5e-41  
 {PIRKEW} magnesium 6e-80  
 {PIRKEW} ATP 2e-97  
 {PIRKEW} polyprotein 1e-34  
 {PIRKEW} alternative initiators 2e-31  
 {PIRKEW} phosphoprotein 2e-74  
 {PIRKEW} apoptosis 8e-38  
 {PIRKEW} cGMP binding 4e-33  
 {PIRKEW} glycoprotein 3e-36  
 {PIRKEW} skeletal muscle 8e-38  
 {PIRKEW} protein kinase 2e-50  
 {PIRKEW} testis 5e-41  
 {PIRKEW} cAMP binding 8e-38  
 {PIRKEW} transforming protein 4e-33  
 {PIRKEW} purine nucleotide binding 7e-52  
 {PIRKEW} calcium binding 7e-45  
 {PIRKEW} alternative splicing 5e-42  
 {PIRKEW} P-loop 7e-52  
 {PIRKEW} lipoprotein 8e-38  
 {PIRKEW} proto-oncogene 4e-33  
 {PIRKEW} segmentation 1e-34  
 {PIRKEW} core protein 1e-34



```

(PIRKW]      muscle 8e-38
(PIRKW]      myristylation 8e-38
(PIRKW]      EF hand 7e-45
(PIRKW]      cell division 3e-49
(PIRKW]      homodimer 1e-32
(PIRKW]      calmodulin binding 5e-42
(SUPFAM]     ribosomal protein S6 kinase II 1e-34
(SUPFAM]     calcium-dependent protein kinase 7e-45
(SUPFAM]     AMP-activated protein kinase 6e-80
(SUPFAM]     protein kinase akt 3e-36
(SUPFAM]     protein kinase SPK1 7e-41
(SUPFAM]     unassigned Ser/Thr or Tyr-specific protein kinases 8e-99
(SUPFAM]     Ca2+/calmodulin-dependent protein kinase 5e-42
(SUPFAM]     calmodulin repeat homology 7e-45
(SUPFAM]     cAMP receptor protein cyclic nucleotide-binding domain homology 3e-33
(SUPFAM]     protein kinase DUN1 6e-36
(SUPFAM]     protein kinase C zeta 4e-33
(SUPFAM]     Dictyostelium cAMP-dependent protein kinase catalytic chain 2e-34
(SUPFAM]     death-associated protein kinase 8e-38
(SUPFAM]     pleckstrin repeat homology 3e-36
(SUPFAM]     ankyrin repeat homology 8e-38
(SUPFAM]     protein kinase homology 8e-99
(SUPFAM]     Ca2+/calmodulin-dependent protein kinase II 6e-38
(SUPFAM]     protein kinase C zinc-binding repeat homology 4e-33
(SUPFAM]     protein kinase C delta 2e-32
(SUPFAM]     cGMP-dependent protein kinase 3e-33
(SUPFAM]     protein kinase cdrl 1e-45
(SUPFAM]     kinase-related transforming protein 2e-50
(SUPFAM]     Ca2+/calmodulin-dependent protein kinase I 8e-42
(SUPFAM]     kinase interaction domain homology 7e-41
(SUPFAM]     gag-akt polyprotein 1e-34
(PROSITE]    PROTEIN_KINASE_ATP      1
(PROSITE]    MYRISTYL                 3
(PROSITE]    AMIDATION                 2
(PROSITE]    CAMP_PHOSPHO_SITE         4
(PROSITE]    CK2_PHOSPHO_SITE          15
(PROSITE]    TYR_PHOSPHO_SITE           2
(PROSITE]    PKC_PHOSPHO_SITE           10
(PROSITE]    ASN_GLYCOSYLATION          2
(PROSITE]    PROTEIN_KINASE_ST         1
(PFAM]       Eukaryotic protein kinase domain
(KW]         Irregular
(KW]         3D
(KW]         LOW_COMPLEXITY      12.31 %

```

```

SEQ      MVMADGPRHLQRGPPVRVGFYDIEGLGKGNFAVVKLGRHRITKTEVAIKIIDKSQLDVAVN
SEG      .....
1ctpe    .....EEECTTTEEEEEETTTTEEEEEEEHHHHHHHC

SEQ      LEKIYREVQIMKMLDHPHIKLYQVMETKSMPLYLVTEYAKNGEIFYDLANHGRLNESEAR
SEG      .....
1ctpe    HHHHHHHHHHHHCCCTTTBCCBBBBBETTEEEEEECTTTTBHHHHHHHHHCCCHHHHH

SEQ      RKFQILSAVDYCHGRKIVHRDLKAENLLDNMNIKIADFGNGFFKSGELLATWCGSP
SEG      .....
1ctpe    HHHHHHHHHHHHHHCCCECCCGGGEETTTTCEEECTTTTEETT-TTBC-CCCCCG

SEQ      PYAAPEVFEGQQYEGPQLDIWSMGVVLVVLVCGALPFDGPTLPILRQVLEGRFRIPIYFM
SEG      .....
1ctpe    GGCCHHHHHHCCBC-HHHHHHHHHHHHHHHHCCCTTTTTTHHHHHHHHHHCCCHCTTTT

SEQ      SEDCEHLIRMLVLDPSKRLTIAQIKENKMWLIEVPVQRPVLPQEQENEPSIGEFNEQV
SEG      .....
1ctpe    CHHHHHHHHHTTTTGGGTTTHHHHHHCGG.....

SEQ      LRLMHSGLIDQQKTIESLQNKSYNHFAAIYFLVLRKSHRSSFPVEQRLDGRQRRPSTI
SEG      .....
1ctpe    .....

SEQ      AEQTVAKAQTVGLPVTMHSPPNMLRLSALLPQASNVEAFSPASGCQAEAAFMEEECVDT
SEG      .....
1ctpe    .....

SEQ      PKVNGCLDPVPPVLVRKGCQSLPSNMETSIDEGLETEGEAEEDPAHAFAEFQSTRSGQ
SEG      .....
1ctpe    .....XXXXXXXXXX.....

SEQ      RRHTLSEVTNQLVMPGAGKIFSMNDSPLSDVDSEYDMGSVQRDNLFLDNPSLKDIML
SEG      .....
1ctpe    .....

```

```

SEQ ANQPSFRMTSPFISLRPTNPAMQALSSQKREVNHRSPVSFREGRRASDTSLTQGI VAFRO
SEG .....
lctpE .....

SEQ HLQNLARTKGILELNKVQLLYEQIGPEADPNLAPAPQLQDLASSCPQEEVSQQQESVST
SEG .....XXXXXXXXXXXXXXXXXXXXXXX.....
lctpE .....

SEQ LPASVHPQLSPRQSLETQYLQHRLQKPSLLSKAQNTCQLYCKEPPRSLEQQLQEHRLQOK
SEG .....XXXXXXXXXXXXXXXXXXXXXXX.....
lctpE .....

SEQ RLFLLKQSQQLQAYFNQMQUIAESSYPQPSQQLPLPRQETPPPSQQAPPFSLTQPLSPVLEP
SEG .....XXXXXXXXXXXXXXXXXXXXXXX.....
lctpE .....

SEQ SSEQMQYSPFLSQYQEMQLQPLPSTSGPRAAPPLPTQLQQQPPPPPPPPPPPPPPGGAAPA
SEG .....XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX.....
lctpE .....

SEQ PLQFSYQTCELPSAASPAPDYPTPCQYPVDGAQSDLTGPDGPRSPGLQEPSSSYDPLAL
SEG xxx.....
lctpE .....

SEQ SELPGLFDCEMLDAVDPOHNGYVLVN
SEG .....
lctpE .....

```

## Prosites for DKF2phtes3\_15k11.1

PS00001	115->119	ASN_GLYCOSYLATION	PDOC00001
PS00001	320->324	ASN_GLYCOSYLATION	PDOC00001
PS00004	258->262	CAMP_PHOSPHO_SITE	PDOC00004
PS00004	355->359	CAMP_PHOSPHO_SITE	PDOC00004
PS00004	481->485	CAMP_PHOSPHO_SITE	PDOC00004
PS00004	584->588	CAMP_PHOSPHO_SITE	PDOC00004
PS00005	257->260	PKC_PHOSPHO_SITE	PDOC00005
PS00005	339->342	PKC_PHOSPHO_SITE	PDOC00005
PS00005	420->423	PKC_PHOSPHO_SITE	PDOC00005
PS00005	475->478	PKC_PHOSPHO_SITE	PDOC00005
PS00005	534->537	PKC_PHOSPHO_SITE	PDOC00005
PS00005	545->548	PKC_PHOSPHO_SITE	PDOC00005
PS00005	554->557	PKC_PHOSPHO_SITE	PDOC00005
PS00005	567->570	PKC_PHOSPHO_SITE	PDOC00005
PS00005	579->582	PKC_PHOSPHO_SITE	PDOC00005
PS00005	670->673	PKC_PHOSPHO_SITE	PDOC00005
PS00006	42->46	CK2_PHOSPHO_SITE	PDOC00006
PS00006	54->58	CK2_PHOSPHO_SITE	PDOC00006
PS00006	128->132	CK2_PHOSPHO_SITE	PDOC00006
PS00006	292->296	CK2_PHOSPHO_SITE	PDOC00006
PS00006	359->363	CK2_PHOSPHO_SITE	PDOC00006
PS00006	394->398	CK2_PHOSPHO_SITE	PDOC00006
PS00006	450->454	CK2_PHOSPHO_SITE	PDOC00006
PS00006	458->462	CK2_PHOSPHO_SITE	PDOC00006
PS00006	484->488	CK2_PHOSPHO_SITE	PDOC00006
PS00006	503->507	CK2_PHOSPHO_SITE	PDOC00006
PS00006	515->519	CK2_PHOSPHO_SITE	PDOC00006
PS00006	534->538	CK2_PHOSPHO_SITE	PDOC00006
PS00006	579->583	CK2_PHOSPHO_SITE	PDOC00006
PS00006	878->882	CK2_PHOSPHO_SITE	PDOC00006
PS00006	893->897	CK2_PHOSPHO_SITE	PDOC00006
PS00007	672->680	TYR_PHOSPHO_SITE	PDOC00007
PS00007	100->108	TYR_PHOSPHO_SITE	PDOC00007
PS00008	372->378	MYRISTYL	PDOC00008
PS00008	871->877	MYRISTYL	PDOC00008
PS00008	905->911	MYRISTYL	PDOC00008
PS00009	134->138	AMIDATION	PDOC00009
PS00009	582->586	AMIDATION	PDOC00009
PS00107	26->50	PROTEIN_KINASE_ATP	PDOC00100
PS00108	138->151	PROTEIN_KINASE_ST	PDOC00100

## Pfam for DKF2phtes3\_15k11.1

HMM\_NAME Eukaryotic protein kinase domain

HMM		*YeigRiIGeGsFGtVYKCiWr.TGeIVAiKIIkkzms.....FlREI	
		Y I++++G+G+F++V++++R T +VAIKII+K++++ + RE+	
Query	20	YDIEGTLGKGNFAVVKLGRHRITKTEVAIKIIDKSQLDVNLKIIYREV	68
HMM		qIMRrLnHPNIIRFYDwFedddDHIYMIMEYMeGGDLFDYIrrngpMsEw	
		QIM++L+HP+II++Y ++E +++ +Y+++EY+ +G++FDY+ ++G+++E	
Query	69	QIMRMLDHPHIKLYQVME-TKSMLYLVTEYAKNGEIFDYLANHGRLNES	117
HMM		eIrfIMyQILrGMeYLHSMgIIHRDLKPENILIDeNgqIKIcDFGLARqM	
		E+R+ ++QIL++++Y+H ++I+HRDLK+EN+L+D+N++IKI+DFG+ ++	
Query	118	EARRKFWQILSAVDYCHGRKIVHRDLKAENLLDNNMNIKIADFGFGNFF	167
HMM		nnYerMtrfCGTPWYMMAPEVImg.nyYttkVDMWSFGCILWEMMTGep	
		+++E++ T CG+P+Y APEV +G +Y +++ D+WS+G++L+ +++G +	
Query	168	KSGELLATWCGSPPYA-APEV-FEGQQYEGPQLDIWSMGVLYVLVCGAL	215
HMM		PFyddnMemImrIiqrfrpfWpnCSeElyDFMrwCwnyDPeKRPTrQI	
		PF++ ++ + + +++ R+++++ +SE++ ++R+++ +DP+KR+T+ QI	
Query	216	PFDGPTLPILRQRVLEGRFRIPYFMSEDCHELIIRRLVLDPSKRLTIAQI	265
HMM		LnHPWF*	
		+H W+	
Query	266	KEHKWM 271	

DKF2phtes3\_17f10

group: testes derived

DKF2phtes3\_15j18 encodes a novel 710 amino acid protein with weak similarity to neurofilament proteins.

No informative BLAST results; No predictive prosite, pfam or SCOP motif.

The new protein can find application in studying the expression profile of testis-specific genes.

similarity to neurofilament proteins

Sequenced by GBF

Locus: unknown

Insert length: 2533 bp

Poly A stretch at pos. 2507, no polyadenylation signal found

```
1  CTTCAAGTTCA ACTAAAAATG GACAGATCTC AGCAGACCAG CCGTACAGGA
51  TACTGGACCA TGATGAACAT CCCCCCTGTA GAAAAAGTGG ACAAGGAACA
101 ACAGACATAC TTTAGTGAAT CAGAAATAGT GGTATTTCCT AGGCCAGATA
151 GTTCTTCTAC AAAGTCAAAG GAAGATGCCG TGAAACATAA ATCTCGGGA
201 AAGATTTTGT CTAGTGAACA CCCTGAATTT CAACCAGCAA CAACAGCAA
251 TGAAGAAATG GGGCAGAAAA ATATCAGCAG AACTTCATTT ACTCAGGAGA
301 CTAAAAAAGG TCCCCAGTA CTTTGAAGAG ATGAGCTTAG GGAAGAAATA
351 ACTGTACCTG TTGTACAAGA AGGTTCTGCT GTTAAAAAAG TGGCTTCTGC
401 TGAATATAGC CCTCCATCAA CAGAAAATT CCCAGCTAAA ATACAGCCTC
451 CATTAGTTGA AGAGGCCACT GCTAAAGCGG AGCCCAGACC TGCTGAAGAG
501 ACCCATGTCC AAGTACAGCC ATCAACTGAA GAGACTCCTG ATGCTGAGGC
551 AGCCACTGCA GTTGCAGAGA ATTCTGTAA AGTTCAGCCT CCACCTGCTG
601 AAGAGGCCCC TTTAGTGGAG TTTCTGCTG AAATTCAGCC TCCTATCAGT
651 GAAGAGTCTC CTCTGTAGA GCTTCTGGCT GAAATTCGTC CTCCATCAGC
701 TGAAGAGTCC CTTTCAAGAG AGCCTCCTCG TGAATTCCTG CCTCCACCAG
751 CTGAAAAATG TCCTTCAGTA GAGCTTCTTG GTGAAATTCG GTCTCCCTCA
801 GCACAAAAGG CTCCCATTTA AGTACAGCCT TTACAGCTG AGGGGCCCTT
851 TGAAGAGGCC CCAGCTAAAG TAGAGCCTCC CACTGTTGAA GAGACCTTGT
901 CTGAAGTTCA GCCTCTATTA CCTGAAGAGG CTCTAGAGA AGAGGCTCGA
951 GAACTTCAGC TTTCAACAGC TATGGAGACC CCTGCAGAAG AGGCTCCTAC
1001 TGAATTTTCA TCTCCATTAC CTAAGAGAGC CACTGCAGAA GAGGCTCTGT
1051 CTGAAATTTA GCTTCTAGCA GCTACGGAGC CTCTGCAGA TGAATCTCCT
1101 GCCGAGGCTC GGTCTCCACT ATCTGAGGAG ACTTCTGAGC AAGAGGCTCA
1151 TGCTGAAGTT CAATCTCCAT TAGCTGAAGA GACCACTGCA GAAGAGGCTT
1201 CTGCTGAAAT TCAGCTTCTA GCAGCTATAG AGGCTCCTCG AGATGAAACT
1251 CCTGCTGAAG CTCAGTCTCC ACTATCTGAG GAGACTTCTG CAGAAGAGGC
1301 TCCTGCTGAA GTTCAGTCTC CATCAGCTAA GGGAGTTTCT ATAGAAGAGG
1351 CCCCCTTTGA GCTTCAGCCT CCATCAGGTG AAGAGACCAC TGCAGAAGAG
1401 GCCTCTGCTG CAATTCAGCT TCTAGCAGCT ACAGAGGCTT CTGCAGAAGA
1451 GGCTCCTGCT GAAGTTCAAG CTCCACCAGC TGAGGAGGCC CCGCTGAAG
1501 TTCAGCCTCC ACCAGCTGAG GAGGCCCCCG CTGAAGTTCA GCCTCCACCA
1551 GCTGAGGAGG CCCCCTGTA AGTTCAGCCT CCACCAGCTG AGGAGGCCCC
1601 CGCTGAAGTT CAGCCTCCAC CAGCTGAGGA GGGCCCCGCT GAAGTTCAAG
1651 CTCCACCAGC TGAGGAGGCC CCCCCTGAAG TTCAGCCTCC ACCAGCTGAG
1701 GAGGCCCTCG CTGAAGTTCA GTCTCTACCA GCTGAGGAGA CTCCATATAG
1751 AGAGACCCTT GCTGCACTAC ACTCTCCCCC AGCTGATGAT GTCCCTGCAG
1801 AAGAGGCTCT CGTTGACAAA CATTCCTCC CAGCTGATTT GCTTCTGACT
1851 GAGGAGTTTC CTATAGGAGA GGCCTCTGCT GAAGTTTAC CTCCACCATC
1901 TGAACAAACC CCTGAAGATG AGGCTCTGTT AGAAGATGTG TCTACAGAAT
1951 TTCAGTCAAC GCAGGTGGCA GGAATTCAG CAGTAAATTT AGGATCGGTT
2001 GTTTTGAAG GTGAAGCAA ATTTGAAGAG GTTTCAAAAA TCAATCTGT
2051 CCTTAAAGAT TTGTCTAATA CCAATGATGG ACAGGCTCCC ACTCTTGAAA
2101 TAGAAAGTGT TTTTCATATA GAATTAAGAC AACGCTCTCC TGAAGTGTAG
2151 TCAGGTTGTA CCTAAGCTAG CAATCAGAAG CTACATGGTT TTGGAAGAAC
2201 ATACTTTAGA AAAGGTTGGG CAGCAGGAAG TAGCTTTGTC AATAAGGCAA
2251 ATTAAGGGG ACCCAAGAC TTGGAATACA GGTGGAATA TGAACATAA
2301 AAATGTAGC AGCATAAAAT TACTGTGTTT AATTTCATTC AAATTTATGG
2351 CATGAAAAAT ACCTATTTTG AAAGTAAAGT TATAATTGAA AAAAATTGCT
2401 TAAATATCC TTTCTTGTCT AAACCTGTGT ACACAGTAA AGTTTAACTT
2451 GCAGCCATCT TTTCTTGTCT TTGCCTTCCC TTTATAAGTA AATATAGTTT
2501 CTAGTGGAAA AAAAAAAAAA AAAAAAAAAA AAA
```

BLAST Results

## No BLAST result

## Medline entries

No Medline entry

## Peptide information for frame 3

ORF from 18 bp to 2147 bp; peptide length: 710  
Category: similarity to known protein  
Classification: unclassified

1 MDRSQQTSTRT GYWTMMNIPP VEKVDKEQQT YFSESEIVVI SRPDSSTKS  
51 KEDALKHKSS GKIFASEHPE FQPATNSNEE IGQKNISRTS FTQETKKGPP  
101 VLLEDELREE VTHVQVQEGS AVKKVASAEI EPPSTEKFFPA KIQQPLVEEA  
151 TAKAEPRPAE ETHVQVQST EETPDAAEAT AVAENSVKVQ PPPAEAPLV  
201 EFPAEIQPPS AEESPSVELL AEILPPSAEE SPSEEPPEAI LPPPAEKSPS  
251 VELLGEIRSP SAQKAPIEVQ PLPAEGALEE APAKVEPPTV EETLAEVQPL  
301 LPEEAPREEA RELQLSTAME TPAEEAPTEF QSPLPKETTA EEASAEIQLL  
351 AATEPPADET PAEARSPLSE ETSAAEAHAE VQSPLAEETT EEASAEIQL  
401 LAIEAPADE TPAEAQSPLS EETSAAEAPA EVQSPSAGV SIEEAPLELQ  
451 PPSGEETTAE EASAAIQLLA ATEASAEAP AEVQPPPAEE APAEVQPPPA  
501 EEAPEVQPP PAEEAPAEVQ PPPAEAPAE VQPPPAEAP AEVQPPPAEE  
551 APSEVQPPPA EEAPEVQSL PAEETPIEET LAAVHSPPAD DVPAAEASVD  
601 KHSPPADLLL TEEFPIGEAS AEVSPPPEEQ TPEDEALVEN VSTEFQSPQV  
651 AGIPAVKLGS VVLEGEAKFE EVSKINSVLK DLSNTNDGQA PTLEIESVFH  
701 IELKQRPEEL

## BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKF2phtes3\_17f10, frame 3

PIR:A37221 neurofilament triplet H protein - rat, N = 1, Score = 480, P  
= 7.4e-43

TRMBL:RNNFLH\_1 Rat heavy neurofilament subunit (NF-H) mRNA, 3' end., N  
= 1, Score = 475, P = 1e-42

>PIR:A37221 neurofilament triplet H protein - rat  
Length = 1,072

## HSPs:

Score = 480 (72.0 bits), Expect = 7.4e-43, P = 7.4e-43  
Identities = 185/622 (29%), Positives = 320/622 (51%)

Query: 33 SESEIVVISRPDSSSTKSKEDALKHKSSGKIFASEHPEFQPATNSNEEIGQKNISRTSFT 92  
SE +I V+ + + + +E + + + + + E E Q E G + + TS  
Sbjct: 436 SEEKIKVVEKSEKETVIVEEQTEETQVTEEVTEEDKEAQGEEEEAEEGGEEAATTSP 495

Query: 93 QETKKGPPVLLEDELREEVTVVQEGSAVKKVASAEIEPPSTEKFFPAKIQQPLVEEATA 152  
E P + ++EE P + A K + AE + P+ K PA+++ P ++ A  
Sbjct: 496 AEEAASPEKETKSPVKEEAKSPAEAKSPAEAK-SPAEAKSPAEVKSPEVKSPEAKSPA 554

Query: 153 KAEPRPAEETHVQVQSTETPDAAEATAVAENSVKVQPPPAEEAP-LVEFPAEIQPPSA 211  
+A+ PAE V+ P+T ++P + A A++ +V+ P ++P + PAE + P+  
Sbjct: 555 EAKS-PAE---VK-SPATVKSPAEAKSPAEAKSPAEVKSPEVKSPEAKSPA 609

Query: 212 EESP-SVELLAELPPSAEESPSE-EPPAEILPPPAEKSPS-VELLGEIRSPSAQKAPIE 268  
+SP + AE P++ +SP E + PAE P KSP+ V+ E +SP+ K+P+  
Sbjct: 610 VKSPVEAKSPAEAKSPASVKSPGEAKSPAEAKSPAEVKSPEVKSPEAKSPA 669

Query: 269 VQPLPAEGALEEAPAKVEPPTVEETLAEVQPLPEEAPREEARELQLSTAMETPAE-EAP 327  
V+ PAE ++P +V+ P ++ +E + ++P E A+ ++PAE ++P  
Sbjct: 670 VKS-PAEA---KSPVEVKSASVKSPSEAKSPAGAKSPA-EAKS---PVVAKSPAEAKSP 721

Query: 328 TEFQSPFLPKETTAEASAEIQLLAATEPPAD-ETPAEARSPLSEETSAAEAHAEVQS--- 383  
E + P ++ AE S A + PA+ ++PAE+SP+ E S E+A + V+  
Sbjct: 722 AEAKPPAEAKSPAEAKSP-----AEAKSPAEAKSPAEAKSPV-EVKSPEKAKSPVKEGAK 775

Query: 384 PLAEETTAEEASAEIQLLAIEAPAD-ETPAEAQSPLEET-SAEAPAE-EVQSPSAGV 440

LAE + E+A + ++ I+ PA+ ++P +A+SP+ EE S E+A +V+SP AK  
 Sbjct: 776 SLAEAKSPEKAKSPVK--EEIKPPAEVKSPEKAKSPMKEEAKSPEKAKTLDVKSPEAKTP 833

Query: 441 SIEEA--PLELOPPSGEETTA--EASAAIQLLAATEASA---EEAPAEVQPPPAEEAPAE 494  
 + EEA P +++ P ++ A EEA + + TE A EE + V+ A+E P +  
 Sbjct: 834 AKEEAKRPADIRSPQVKSPEAKEAKSPEKEETRTKVPAPKKEEVKSPVEEVKAKEPPKK 893

Query: 495 VQPPPAEEAP-AEVQPPPAEEAPAEVQPPPAEEAPAEVQPPPAEEAPAEVQPPPAEEAPS 553  
 V+ P EV+ +EAP E Q P AEE + P +++P E + EEA  
 Sbjct: 894 VEEETPATPKTEVKESKKDEAPKEAQKPAEEKEPLTEKP--KDSPEGAKK---EEAKE 948

Query: 554 EVQPPPAEEAPAEV---QSLP---AEETPIEETL--AAVHSPADDVPAEEASVD-KHS 603  
 + P EE PA++ ' ++ P AE+ +E + P ++VPA D K  
 Sbjct: 949 KKAAPAEETPAKLGVKKEAKPKEAKEDAKAKEPSKPSKEKPKKEEVPAPEKKDTEKE 1008

Query: 604 PPADLLTTEEFPIGEASAEVSPP--PSEQT-PEDEALVENVSTEFQSPQ 649  
 + EE P +A A+ P E + P+ E ++ ST+ + Q  
 Sbjct: 1009 KTTESKKPEEKPMQAKAKEEDKGLPQEPSKPKTEKAESSSTQKDSQ 1057

Score = 473 (71.0 bits), Expect = 4.8e-42, P = 4.8e-42  
 Identities = 184/628 (29%), Positives = 310/628 (49%)

Query: 18 IPPVEKVDKEQTYFSESEIVVISRP---DSSSTKSKEDALKHKSSGKIFASEHPEFQPA 74  
 I VEK +KE ++E + ++ + E+ + + G+ A+ P + A  
 Sbjct: 440 IKVVEKSEKETVIVEEQTEEQTEEVTEEDKEAQGESEEEAEEGGEAAATSPPAEEA 499

Query: 75 TNSNEEIQKNISRTSFTQETKKGPPVLEDELREEVTVVQEGSAVKKVASAEIEPPS 134  
 + +E + + + + K P E + E P + A K + AE + P+  
 Sbjct: 500 ASPEKET-KSPVKEEAKSPAEEAKSPA---EAKSPAEEAKSPAEEAKSPAEEAKSPA 554

Query: 135 TEKFPKIQPPLVEEATAKAEPRAETHVQVQ-PSTEETPDAAEATAVAENSVKVPQPP 193  
 K PA+++ P ++ A+A+ ++ +V+ P+T ++P + A A++ +V+ P  
 Sbjct: 555 EAKSPAEEAKSPAEEAKSPAEEAKSPAEEAKSPAEEAKSPAEEAKSPAEEAKSPA 614

Query: 194 AEEAPL-VEFPAEIQPPSAEESPS-VELLAELPPSAEESPS-EPPAEILPPPAEKSPS 250  
 ++P + PA ++ P +SP+ + AE+ P+ +SP E + PAE+ P KSP+  
 Sbjct: 615 EAKSPAEEAKSPAEEAKSPAEEAKSPAEEAKSPAEEAKSPAEEAKSPAEEAKSPA 674

Query: 251 -VELLGEIRSPSAQKAPIEVQ-PLPAEGALE-EAPAKVEPPTVEETLAEVQPLLPPEAPR 307  
 + E+SP++ K+P E + P A+ E ++P + P ++ AE + P ++P  
 Sbjct: 675 EAKSPAEEAKSPAEEAKSPAEEAKSPAEEAKSPAEEAKSPAEEAKSPAEEAKSPA 734

Query: 308 EEARELQSTAME--TPAE-EAPTEFQSP----LP-KE---TTAEASAEIQLLAATE-- 354  
 E + + E +PAE ++P E +SP P KE + AE S E E  
 Sbjct: 735 EAKSPAEEAKSPAEEAKSPAEEAKSPAEEAKSPAEEAKSPAEEAKSPAEEAKSPA 794

Query: 355 -PPAD-ETPAEARSPLSEET-SAEAAH-EVQSPLEAETTAEAS--AEIQLLAATEAPA 408  
 PPA+ ++P +A+SP+ EE S E+A +V+SP A+ EEA A+I+ +++PA  
 Sbjct: 795 KPPAEVKSPEKAKSPMKEEAKSPEKAKTLDVKSPEAKTPEAKRPAADIRSPQVKS 854

Query: 409 DETPAEQSPLSEETSAAE-APA--EVQSPSAKGVSEIAPLELOPPSGEETTAEASAA 465  
 E EA+SP EET E+ AP EV+SP +EE + +PP E EE + A  
 Sbjct: 855 KE---EAKSPEKEETRTKVPAPKKEEVKSP-----VEEVKAK-EPPKKVE---EETPA 901

Query: 466 IQLLAATEASAEAPAEVQPPPAEEAPAEVQPPPAEEAPAEVQPPPAEEAPAEVQPPPAE 525  
 E+ +EAP E Q P AEE + P +++P E + A+E A P E  
 Sbjct: 902 TPKEVKESKKDEAPKEAQKPAEEKEPLTEKP--KDSPEGAKEEAKKAAA---PEE 956

Query: 526 EAPAEV---QPPPAEEAPAEVQPPPAEEAPSEVQPPPAEEAPAEVQSLPAEETPIEETL 581  
 E PA++ + P E+A P++ PSE + P EE PA + +E E+  
 Sbjct: 957 ETPAKLGVKKEAKPKEAKEDAKAKEPSK--PSEKEKPKKEEVPAPEKKDTEETTESK 1014

Query: 582 AAVHSPADDVPAEEASVDKHSPPADLL-LTEEFPIGEASAEVSPPSPSEQTPDEEA 636  
 P EE DK P TE+ ++ + PSE+ PED+A  
 Sbjct: 1015 KPEEKPMQAKAKEE---DKGLPQEPSKPKTEKAESSSTQKDSQSEKAPEDKA 1067

Score = 421 (63.2 bits), Expect = 3.7e-36, P = 3.7e-36  
 Identities = 162/540 (30%), Positives = 275/540 (50%)

Query: 135 TEKFPKIQPPLVEEATAKAEPRAETHVQVQSTEETPDAAEATAVAENSVKV 189  
 TE P KI P + K+E + +E+ V V+ TEE E T E +  
 Sbjct: 419 TEGLP-KI-PSMSTHIKVKSEKIKVVEKSEKETVIVEEQTEEQTEEVTE--EEDKEA 474

Query: 190 QPPPAEEAPLVEFPAEIQPPSAEESPSVELLAELPPSAEE--SPSE-EPPAEILPPPAE 246  
 Q EEA A P AEE+ S E E P EE SP+E + PAE P  
 Sbjct: 475 QGEEEEEAEEGGEAAATSPPAEEAASPE--KETKSPVKEEAKSPAEEAKSPAEEAKSPA 532

Query: 247 KSPSVELLGEIRSPSAQKAPIEVQPLPAEGALEEAPAKVEPPTVEETLAEVQPLLPPEAP 306  
 KSP+ E++SP+ K+P E + PAE ++PA+V+ P ++ AE + +P  
 Sbjct: 533 KSPA-----EVKSPAEEAKSPAEEAKSPAEEAKSPAEEAKSPAEEAKSPA 583

Score = 406 (60.9 bits), Expect = 1.7e-34, P = 1.7e-34  
Identities = 123/390 (31%), Positives = 213/390 (54%)

Score = 255 (38.3 bits), Expect = 5.5e-18, P = 5.5e-18  
Identities = 124/420 (29%), Positives = 199/420 (47%)

615

Query:	525	EEAPAEQPPPAEAPAEVQPPPAEAEAEVQPPPAEAEAPAEVQSLPAE-ETPIE-ETLA	582
		++PAE + P ++PAEV + ++PE + P ++PA V+S PAE ++P E ++ A	
Sbjct:	526	AKSPA EAKSPAEVKS PAEVKSPAEAKSPAEAKSPAEVKS PATVKS -PAEAKSPAEAKSPA	584
Query:	583	AVHSPAPDDVPADVPAAEKVDKHSPPADLLTEEFIPGEASAEVSPSPSEQT-P-EDEALVENV	641
		V SP P E S + PA++ ++ AE P S ++P E ++ E	
Sbjct:	585	EVKSPATPVKSPGEAKSPAEAKSPAEVKS PVE--AKSPA EAKSPASVKSPGEAKSPAEAK	641
Query:	642	S-TEFOSPOVAGIP	654
		S E +SP P	
Sbjct:	642	SPA EVKSPATPVKS P	655

Score = 253 (38.0 bits), Expect = 9.0e-18, P = 9.0e-18  
Identities = 115/364 (31%), Positives = 166/364 (45%)

Query:	110	EVTVPVVOEGSAVKKVASAEIEPPSTEFKPAKIQPLVEEATAKAEPRAE-ETHVQVO-	167
		ES PVV + A K + AE + PP+AK PA+ + P + +A+A+ PAE + +V+	
Sbjct:	705	EAKSPVKGAKSPAEAK-SPAEEKPPAEAKSPAEAKSPAEAKSPAEAKSPVEVKS	762
Query:	168	PSTEETPDAAEAATAVAE--NSVKVQPPPAEEA--PL-VEFPAAIQPPSAEE--SPSVELL	220
		P ++P EA A++AE + K + P EE P V+ P + + P EE SP	
Sbjct:	763	PEKAKSPVKGAKSLAEAKSPKAKSPVKEEIKPPAEVKSPEKAKSPMKEAKSPKAKT	822
Query:	221	AEILPPSAEESPEEP--PAEILPPPAEKSPSVELLGEIRSPSAQKAPIE-VQPLPAE--	275
		+ + P+A+ EE PA+I P KSP+ E E +SP + + + V P E	
Sbjct:	823	LDVKSPEAKTPAKEAEKRPADIRSPQEVKSPAKE--EAKSPKEETRTREKVPKKEEVK	879
Query:	276	GALEEAPAKVPEPTVEETLAEVQPLLPEAPREEAREIQLSAMETPAEEA-P-TFQSP	333
		+EE AK P VEE E P P+ +E+K + A + AEE P TE	
Sbjct:	880	SPVEEAKPEKPPKVEE--EKTPTATPK+VEKSSKDOEAPKQAPKPEKEPTLTKPKD	936
Query:	334	LPKETTAEASAEIQLLAATEPPADETPAE--ARSPLEETSAEEAHA-EVQSPLAETT	390
		P E EEA + AA P +ETPA+ + + AE+A A E P + E	
Sbjct:	937	SPGEAKKEAEK--KAAA--PEEETPAKGVKEAKPEKAEADAKPEKSPKSEKPK	991
Query:	391	A-EASAEIQLLAIEAPADETPAEAQSPLEETSAEEAPAEVQSPSA-KGVSIEEAPLE	448
		EE A + E E+ + P + + EE Q PS K E++	
Sbjct:	992	KKEEVPAPEKKDKTEKEETSSKKPEKPKMQAKAEEDKGLPQEPSPKPKTAEKSSST	105
Query:	449	LQPPSGEETTAEEASAA 465	
		Q S A E AA	
Sbjct:	1052	DQKDSQSEKAPEDKAA 1068	

Pedant information for DKF2phtes3\_17f10, frame 3

Report for DKFZphtes3 17f10.3

```
[LENGTH]      710
[MW]           75131.94
[pI]           4.02
[KW]           All_Alpha
[KW]           LOW_COMPLEXITY      34.08 %
```

```
SEQ      MDRSQQTSRTGYWTHMNIIPPEVKVDKEQQTYFSESEIVVISRPDSSSTKSKEDALKHKSS
SEG
PRD      cccccccccccccccccccccceehhhhhhhccccceeeccccccccchhhhhhhhhccc

SEQ      GKI FASEHPFQPATNSNEEI GQKNI SRSTS FTQETKKGPPVLLEDELREEVTVPVVQEGS
SEG
PRD      cceecccccccccccccccccccccccccceeeccccchhhhhhhhhhheeeccccc

SEQ      AVKKVASAEIEPPSTEKFPKAIQPLLVEEATAKAEPRAEETHVQQPSTETPDAAEAT
SEG
PRD      chhhhhhhccccccccccccccccchhhhhhhhhccccceeeccccccccchhhhh

SEQ      AXNENSKVQPPPAEAPLVEFPAEI QPPSAEESPVELLAEILPPSAEESPPEPAEI
SEG
PRD      hhhhccccccccccccceeeccccccccccccccccchhhhhhhcccccccccccccccc

SEQ      LPPPAEKSPSVELLGEIRSPSAQKAPIEVQPLPAEGALEAPAKVEPTVEETLAEVQPL
SEG
PRD      xxxxxxxx          xxxxxxxxxxxxxxxx          xxx

SEQ      LPPEAPREAREQLQSTAMETPAEAPETFQSPPLKPTTAEASAEIQLLAETPPADET
SEG
PRD      xxxxxxxxxx          xxxxxxxxxx          xxxxxxxxxx
ccccchhhhhhhhhhhhhhhccccccccccccccccchhhhhhhhhhhhhhhhhcccccc
```



(No Prosite data available for DKFZphtes3\_17f10.3)  
(No Pfam data available for DKFZphtes3\_17f10.3)

DKF2phtes3\_17117

group: metabolism

DKF2phtes3\_17117 encodes a novel 626 amino acid protein with similarity to transketolases (EC 2.2.1.1).

The novel protein contains a ATP/GTP-binding site motif A (P-loop). It is a new testis-specific transketolase. Transketolase requires thiamin pyrophosphate as cofactor and shows a wide specificity for both reactants, e.g. converts hydroxypyruvate and R-CHO into CO(2) and R-CHOH-CO-CH(2)OH.

The new protein can find application in modulation of metabolic pathways involving this transketolase activity and as a new enzyme for biotechnologic production processes.

strong similarity to transketolases

few EST hits (all from testis or pooled librarys containing testis)  
testis specific transketolase?

Sequenced by GBF

Locus: unknown

Insert length: 2688 bp

Poly A stretch at pos. 2649, polyadenylation signal at pos. 2630

```
1  GACAAAAGAG AGATGATGGC CAACGACGCC AAGCCCGACG TGAAGACCGT
51  GCAGGTGCTG CGGGACACAG CCAACCGCCT GCGGATCCAT TCCATCAGGG
101 CCACGTGTGC CTCTGGTTCT GGCCAGCTCA CGTCGTGCTG CAGTGCAGCG
151 GAGGTGCTGT CTGTCTCTCT CTCCACACG ATGAAGTATA AACAGACAGA
201 CCCAGAACAC CCGGACACAG ACCGGTTTAT CCTCTCCAGG GGACATGCTG
251 CTCCTATCCT CTATGCTGCT TGGGTGGAGG TGGGTGACAT CAGTGAATCT
301 GACTTGCTGA ACCTGAGGAA ACTTCACAGC GACTTGAGAA GACACCCCTAC
351 CCCGCGATTG CCGTTTGTGT ACGTGGCAAC AGGGTCCCTA GGTCAAGGAT
401 TAGGTACTGC ATGTGGAATG GCTTATACGT GCAAGTACCT TGACAAGGCC
451 AGCTACCGGG TGTCTGCTCT TATGGGAGAT GGCGAATCCT CAGAAGGCTC
501 TGTGTGGGAG GCTTTTGTCT TTGCCTCCCA CTACAACCTG GACAATCTCG
551 TGGCGGTCTT CGACGTGAAC CGCTTGGGAC AAAGTGGCCC TGCACCCCTT
601 GAGCATGGCG CAGACATCTA CCAGAATTGC TGTGAAGCCT TTGGATGGAA
651 TACTTACTTA GTGGATGGCC ATGATGTGGA GGCCTTGTGC CAAGCATTTT
701 GGCAAGCAAG TCAAGTGAAG AACAAAGCCTA CTGCTATAGT TGCCAAGACC
751 TTCAAAGGTC GGGGTATTCC AAATATTGAG GATGCAGAAA ATTGGCATGG
801 AAAGCCAGTG CCAAAAGAAA GAGCAGATGC AATTGTCAAA TTAATTGAGA
851 GTCAGATACA GACCAATGAG AATCTCATAC CAAAATCGCC TGTGGAAGAC
901 TCACCTCAAA TAAGCATCAC AGATATAAAA ATGACCTCCC CACCTGCTTA
951 CAAAGTTGGT GACAAGATAG CTAATCAGAA AACATATGGT TTGGCTCTGG
1001 CTAAGCTGGG CCGTGCAAAAT GAAAGAGTTA TTGTTCTGAG TGGTGACACG
1051 ATGAACTCCA CTTTTCTGA GATATTACAG AAAGAACACC CTGAGCGTTT
1101 CATAGAGTGT ATTATTGCTG AACAAAACAT GGTAAAGTGT GCACTAGGCT
1151 GTGCTACACG TGGTCGAACC ATTGCTTTTG CTGGTGCTTT TGCTGCCTTT
1201 TTTACTAGAG CATTGATGCA GCTCCGAATG GGAGCCATTG CTCAAGCCAA
1251 TATCAACCTT ATTGGTTCCC ACTGTGGGGT ATCCACTGGA GAAGATGGAG
1301 TCTCCAGATG GGCCCTGGAG GATCTAGCCA TGTCCGAAG CATTCCCAAT
1351 TGTACTGTTT TCTATCCAAG TGATGCCATC TCGACAGAGC ATGCTATTTA
1401 TCTAGCCGCC AATACCAAGG GAATGTGCTT CATTGGAACC AGCCAAACAG
1451 AAACCTGCAGT TATTTATACC CCACAAGAAA ATTTTGAGAT TGGCCAGGCC
1501 AAGGTGGTCC GCCACGGTGT CAATGATAAA GTCACAGTAA TTGGAGCTGG
1551 AGTTACTCTC CATGAAGCCT TAGAAGCTGC TGACCATCTT TCTCAACAAG
1601 GTATTTCTGT CCGTGTATC GACCAATTTA CCATTAAACC CCTGGATGCC
1651 GCCACCATCA TCTCCAGTGC AAAAGCCACA GGCGCCGAG TTATCACAGT
1701 GGAGGATCAC TACAGGGAAG GTGGCATTGG AGAAGCTGTT TGTGCAGCTG
1751 TCTCCAGGGA GCCTGATATC CTTGTTATC AACTGGCAGT GTCAGGAGTG
1801 CCTCAACGTG GGAATACTAG TGAATTGCTG GATATGTTTG GAATCAGTAC
1851 CAGACACATT ATAGCAGCCG TAACACTTAC TTTAATGAAG TAAACTAGGC
1901 TTATTTCTAA AAGTCAAGT CTATTGGCTT TGGCCCAAAA GCACTGGTAT
1951 CTTTGTATTA AATTCATGTT TATTGTCACA AAACCATTAT TTATACCTAT
2001 ACAGTTGTAC TGTTCCTTTT AAAGCAAGC CATTAAACAT CTTTCTTCAT
2051 TCCTAATTTG GAAATTAAG TTTACCTTTC TGTTAATCTA TGTATAAATG
2101 TTACTCTGAG TTATTAATGT GGATTTTAAA ATGTGAAGCA ATAGAATAGG
2151 AAATAAAACA ACTACCTAAT ACAAAATATT CTGATAAGAC TACAAATATC
2201 TGACTGAGCT GGGGATTAAA GTAGAGGTAA CTGTATCTTA AATGAGTATG
2251 ATTTCTCTGT AAGTTAAAAA AATTGAAATT TAATTGTAGA CTTCAATAGT
2301 CCAAGTTTGT AAGGATGTTT GAGCTTTTGT ATAATGCCAT TTATACCTGC
2351 AGTTTACAG ATAAATGTTG ACTGCAGTTG CCTTGGAAAT TCCTCCAAAG
2401 TTTGCCCTCA TCTCTCTCTC ACAGTTTGGG GGTGATGGTG CAGCAGTGGG
2451 ACATCTCTTG ATGCACCACA CTACTTGTGT TCTGTGAAGT GATGAAAGTA
```

2501 TAACTGGTTC TAGTTTGCAC ACTACACACA TAGTTTGTG AAGCTTCAGA  
2551 AATGTTTTTT CTTTTCCTTG TGGCCAAACC AGTTTGTAA TCTGATTATA  
2601 TTCATCTGCT AATGATACTA AAGTTAATGT AATAAGCAT TTAATAATCA  
2651 GAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA

## BLAST Results

No BLAST result

## Medline entries

96214928:  
Amplification of the transketolase gene in desensitization-resistant  
mutant  
Y1 mouse adrenocortical tumor cells.

99123875:  
Properties and functions of the thiamin diphosphate dependent enzyme  
transketolase.

## Peptide information for frame 1

ORF from 13 bp to 1890 bp; peptide length: 626  
Category: strong similarity to known protein  
Classification: Metabolism  
Prosites motifs: ATP\_GTP\_A (595-603)

1 MMANDAKPDV KTVQVLRDTA NRLRIHSIRA TCASGSGQLT SCCSAAEVVS  
51 VLEFFHTMKYK QTDPEHPDND RFILSRGHAA PILYAANVEV GDISESDLLN  
101 LRKLHSDLER HPTPRLPFVD VATGSLGQGL GTACGMAYTG KYLDKASYRV  
151 FCLMGDGESE EGSVWEAFAP ASHYNLDNLV AVEDVNRLLGQ SGPALEHGA  
201 DIYQNCCEAF GWNTYLVDPGH DVEALCQAFW QASQVKNKPT AIVAKTFKGR  
251 GIPNIEDAEN WHGKVPVKER ADAIVKLIES QIQTNENLIP KSPVEDSPQI  
301 SITDIKMTSP PAYKVGDKIA TQKTYGLALA KLGRANERVI VLSGDTMNST  
351 FSEIFRKEHP ERFIECIIAE QNMVSVLALGC ATRGRTIAFA GAFAAFETRA  
401 FQQLRMGAIS QANINLIGSH CGVSTGEDGV SQMALEDLAM FRISIPNCTVF  
451 YPSDAISTEH AIYLAANTKG MCFIRTSQPE TAVIYTPQEN FEIGQAKVVR  
501 HGVNDKVTVI GAGVTLHEAL EAADHLSQQG ISVRVIDPFT IKPLDAATII  
551 SSAKATGGRV ITVEDHYREG GIGEAUCAAV SREPDILVHQ LAVSGVPQRG  
601 KTSLELDMPG ISTRHIIAAV TLTLMK

## BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKF2phtes3\_17l17, frame 1

SWISSPROT:TKT\_MOUSE TRANSKETOLASE (EC 2.2.1.1) (TK) (P68).. N = 1,  
Score = 2222, P = 2.5e-230

SWISSPROT:TKT\_RAT TRANSKETOLASE (EC 2.2.1.1) (TK).. N = 1, Score =  
2202, P = 3.3e-228

TREMBL:RN09256\_1 product: "transketolase"; Rattus norvegicus  
Sprague-Dawley transketolase mRNA, complete cds., N = 1, Score = 2202,  
P = 3.3e-228

SWISSPROT:TKT\_HUMAN TRANSKETOLASE (EC 2.2.1.1) (TK).. N = 1, Score =  
2200, P = 5.3e-228

>SWISSPROT:TKT\_MOUSE TRANSKETOLASE (EC 2.2.1.1) (TK) (P68).  
Length = 623

## HSPs:

Score = 2222 (333.4 bits), Expect = 2.5e-230, P = 2.5e-230  
Identities = 417/614 (67%), Positives = 501/614 (81%)

Query: 7 KPDVKTVQVLRDTANRLRIHSIRATCASGSGQLTSCCSAAEVVSVLEFFHTMKYKQTDPEH 66

Sbjct: KPD + +Q L+DTANRLRI SI+AT A+GSG TSCCSAAE+++VLFHTM+YK DP +  
 6 KPDQKQLQALKDNTANRLRISSIQATTAAGSGHPTSCCSAAEIMAVLFFHTMRYKALDPRN 65

Query: 67 PDNDRFILSRGHAAPILYAAMVEVGDISESDDLNLRLKHSDLERHPTPRLPFVDVATGSL 126  
 P NDRF+LS+GHAAPILYA W E G + E++LLNLRK+ SDL+ HP P+ F DVATGSL

Sbjct: 66 PHNDRFVLSKGHAAPILYAVWAEAGFLPEAELLNLRKISSDLGDHPVPKQAFDQVATGSL 125

Query: 127 GQGLGTACGMAYTGKYLKASYRVFLMGDGESEGSVWEAFASFASHYNLNLVAVFDVN 186  
 GQGLG ACGMAYTGKY DKASYRV+C++GDGE SEGSVWEA AFA Y LDNLVA+FD+N

Sbjct: 126 GQGLGAACGMAYTGKYFDKASYRVYCHLGDGEVSEGSVWEAMAFAGIYKLDNLVAIFDIN 185

Query: 187 RLQSGGPAPLEHGADIYQNCCEAFGWNTYLVGDHDEALCQAFWQASQVKNKPTAIVAKT 246  
 RLQGS PAPL+H DIYQ CEAFGW+T +VDGH VE LC+AF QA K++PTAI+AKT

Sbjct: 186 RLQGS DPAPLQHQVDIYQKRCEAFGWHTIIVDGHVVEELCKAFGQA---KHQPTAIIAKT 242

Query: 247 FKGRGIPNIEDAENWHGKVPKERADAIVKLIESQIQTENLIPKSPVEDSPQISITDIK 306  
 FKGRGI IED E WHGKP+PK A+ I++ I SQ+Q+ + ++ P ED+P + I +I+

Sbjct: 243 FKGRGITGIEDKEAWHGKPLPKNMAEQIIQEIYSQVQSKKILATPPQEDAPSVDIANIR 302

Query: 307 MTPSPAYKVGDKIATQRTYGLALAKLGRANERVIVSGDTMNSTFSEIFRKEHPERFIEC 366  
 M +PP+YKVGDKIAT+K YGLALAKLG A++R+I L GDT NSTFSE+FKHP+RFIEC

Sbjct: 303 MTPSPSYKVGDKIATRKAYGLALAKLGHASDRIIALDGTKNSTFSEIFRKEHPRDFIEC 362

Query: 367 IIAEQNMVSVAGCATRGRTIAFAGFAFAFFTRAFDQLRMGAISQANINLIGSHCGVSTG 426  
 IAEQNMVS+A+GCATR RT+ F FAFFTRAFDQ+RM AIS++NINL GSHCGVS G

Sbjct: 363 YIAEQNMVSIAGCATRDRTVPFCSTFAFFTRAFDQIRMAAISNINLIGSHCGVSGIS 422

Query: 427 EDGVSQMALEDLAFMRSPNCTVFYPSDAISTEHAIIYLAANTKGMCFIRTSQPETAIVIY 486  
 EDG SQMALEDLAFMRSP TVFYPSD ++TE A+ LAANTKG+CFIRTS+PE A+IY+

Sbjct: 423 EDGVSQMALEDLAFMRSPVPMSTVFYPSDGVATEKAVELAANTKGCFIRTSRPEAIYIS 482

Query: 487 PQENFEIQQAQVVRHGVNDKVTIVIGAGVTLHEALEAADHLSQGGISVRVIDPFTIRPLDA 546  
 E+F++GQAQVV +D+VTIVIGAGVTLHEAL AA+ L + IS+RV+DPFTIRPLD

Sbjct: 483 NNEDFQVQAQVVLKSKDDQVTIVIGAGVTLHEALAAEELKKDKISIRVLDPFTIRPLDR 542

Query: 547 ATIISAKATGGRVITVEDHYREGGIGEAACAASREPDIIVHQLAVSGVPQRGKTSSELL 606  
 I+ SA+AT GR++TVEDHY EGGIGEA AAV EP + V +LAVS VP+ GK +ELL

Sbjct: 543 KLILDSARATKGRILTVEDHYEGGIGEAASAAGVGEFVTVTRLAVSQVPRSGKPAELL 602

Query: 607 DMFGISTRHIIAAV 620  
 MFGI I+ AV

Sbjct: 603 KMFGIDKDAIVQAV 616

Pedant information for DKF2phtes3\_17117, frame 1

Report for DKF2phtes3\_17117.1

[LENGTH] 626  
 [MW] 67877.52  
 [PI] 5.90  
 [HOMOL] SWISSPROT:TKT\_MOUSE TRANSKETOLASE (EC 2.2.1.1) (TK) (P68). 0.0  
 [FUNCAT] m outer membrane and cell wall [M. jannaschii, MJ0681] 3e-48  
 [FUNCAT] g carbohydrate metabolism and transport [H. influenzae, HI1023] 9e-36  
 [FUNCAT] 01.05.01 carbohydrate utilization [S. cerevisiae, YPR074c] 5e-32  
 [FUNCAT] 30.03 organization of cytoplasm [S. cerevisiae, YPR074c] 5e-32  
 [FUNCAT] 02.07 pentose-phosphate pathway [S. cerevisiae, YPR074c] 5e-32  
 [FUNCAT] 01.01.01 amino-acid biosynthesis [S. cerevisiae, YPR074c] 5e-32  
 [FUNCAT] i lipid metabolism [H. influenzae, HI1439] 3e-17  
 [FUNCAT] c energy conversion [H. influenzae, HI1233] 2e-09  
 [FUNCAT] 02.01 glycolysis [S. cerevisiae, YBR221c PDB1 - pyruvate dehydrogenase] 2e-05  
 [FUNCAT] 30.16 mitochondrial organization [S. cerevisiae, YBR221c PDB1 - pyruvate dehydrogenase] 2e-05  
 [BLOCKS] BL00801F  
 [BLOCKS] BL00801E  
 [BLOCKS] BL00801D Transketolase proteins  
 [BLOCKS] BL00801C Transketolase proteins  
 [BLOCKS] BL00801B Transketolase proteins  
 [BLOCKS] BL00801A Transketolase proteins  
 [SCOP] dltcrka2 3.28.1.2.1 Transketolase Transketolase, C-terminal domain 1e-21  
 [EC] 1.2.4.1 Pyruvate dehydrogenase (lipoamide) 8e-11  
 [EC] 1.2.4.4 3-Methyl-2-oxobutanoate dehydrogenase (lipoamide) 4e-10  
 [EC] 2.2.1.1 Transketolase 0.0  
 [EC] 2.2.1.3 Formaldehyde transketolase 1e-20  
 [PIRKEW] transferase 0.0  
 [PIRKEW] flavoprotein 2e-07  
 [PIRKEW] Calvin cycle 1e-40  
 [PIRKEW] heterotetramer 2e-07

```

[PIKKW]    pentose phosphate pathway 0.0
[PIKKW]    magnesium 1e-40
[PIKKW]    thiamine pyrophosphate 0.0
[PIKKW]    oxidoreductase 7e-12
[PIKKW]    fatty acid biosynthesis 4e-10
[PIKKW]    mitochondrion 2e-07
[PIKKW]    peroxisome 1e-20
[PIKKW]    homodimer 1e-40
[SUPFAM]    pyruvate dehydrogenase (lipoamide) alpha chain 1e-06
[SUPFAM]    pyruvate dehydrogenase (lipoamide) beta chain 7e-12
[SUPFAM]    ferredoxin 2[4Fe-4S]-related protein 8e-47
[SUPFAM]    thiamine pyrophosphate-binding domain homology 0.0
[SUPFAM]    pyruvate dehydrogenase (lipoamide) 6e-08
[SUPFAM]    ferredoxin 2[4Fe-4S] homology 8e-47
[SUPFAM]    hypothetical protein C2814 2e-21
[SUPFAM]    transketolase 0.0
[PROSITE]   ATP_GTP_A 1
[PFAM]      Transketolase
[KW]        Alpha_Beta
[KW]        3D
[KW]        LOW_COMPLEXITY 3.04 %

```

```

SEQ      MMANDAKPDVKTQVLRDTANRLRIHSIRATCASGSGQLTSCCSAAEVSVLFFHTMKYK
SEG      .....
1ngsB    .....HHHHHHHHHHHHHCCCHHHHHHHHHHHHHHHHHHHHHHHHCCCT

SEQ      QTDPEHPDNDRFILSRGHAAPILYAAWVEVDISESDLNLRKLSDLERHPTPLPFVD
SEG      .....
1ngsB    TTTTTTTTCEEEETTGGGHHHHHHHHHHHCTTCHHHHHHTTTTTTTTTTTTTTTTTTTC

SEQ      VATGSLGQLGTACGMAYTGKYLKASYRVFLMGDGESSEGSVWEAFASFASHYNLNLV
SEG      .....
1ngsB    CCCCTTTHHHHHHHHHHHHHHHHCBTTBTTEEECHHHHHCHHHHHHHHHHHHHHCTTTEE

SEQ      AVFDVNLGQSGPAPLEHGADIYQNCCEAFGWNTYLVGDHVEALCQAFWQASQVKNKPT
SEG      .....
1ngsB    EEEEECCCEETTEEGGGCCCCCHHHHH-HHHCCCEEEETTTTHHHHHHHHHHHHHHTTTTCE

SEQ      AIVANTFKGRGIPNIEDAENWHGKVPKERADAIVKLIESQIQTNENLIPKSPVEDSPQI
SEG      .....
1ngsB    EEEEECTTTTTCCHHHHHHHHHHHTCCHHHHHHHHHHHHHHHHHHHHHHHHHHHHHCHHH

SEQ      SITDIKMTSPPAYKVGDKIATQKTYGI-ALAKLGRANERVIVLSGDTMNSTFSEIFRKEHP
SEG      .....
1ngsB    HHHHHHHHHHTCCCTTTTCBCHHHHHHHHHHHHHHTTTTTTEEEETTTTHHHHCCCTTCECCCG

SEQ      ERFIECIIAEQNMVSVLGCATRGRTIAFAGAFAAFFTRAFDQLRMGAISOANINLIGSH
SEG      .....
1ngsB    GCEETTTTHHHHHHHHHHHHHHTTTTEEEEGGGGGGHHHHHHHHHHHCTTTEEEEC

SEQ      CGVSTGEDGVSQMALEDLAFRSIPNCTVFYPSDAISTEHAIIYLAANTKGMCFFIRTSQPE
SEG      .....
1ngsB    CCGGGTTTTTTTTCCHHHHHHHHCTTTTEEECCCHHHHHHHHHHHHTTTTCEEECCCCCB

SEQ      TAVIYTPQENFEIGQAKVVRHGVNDKVTIVIGAVTLHEALEADHLSQOGISVRVIDPFT
SEG      .....
1ngsB    CCTTTTCHHHHHCC-CEEEETTTTTEEEECCHHHHHHHHHHHHHHHHHHHHCCCEEE...

SEQ      IKPLDAATIISSAKATGGRVITVEDHYREGGIGEAVCAAVSREPDIIVHQLAVSGVPORG
SEG      .....
1ngsB    .....

SEQ      KTSSELLDMFGISTRHIIAAVTLTLMK
SEG      .....
1ngsB    .....

```

## Prosites for DKFZphtes3\_17117.1

PS00017 595->603 ATP\_GTP\_A PDOC00017

## Pfam for DKFZphtes3\_17117.1

HMM\_NAME Transketolase

HMM \*vNtIRiLaMDAVEKANSGHGPGaPMGMAPMAHVLWqrMMRHNPNDPrWPN

Query	20	+N++RI ++ A + +SG +++++A++ VL++++M++++DP P+ ANRLRIHSIRATCASGSGQLTSCCSAAEVSVLFFHTMKYKQTDPEHPD	68
HMM		RDRFVLSNGHacMLLYsMWHLyGYDmPMWDLkQFRQWHSrTPGHPEIghT +DRF+LS GHa+++LY+ W + G +++++DL+++R++HS++ +HP ++	
Query	69	NDRFILSRGHAAPILYAAWVEVGD-ISESDLNLRKLHSDLERHPTPLRP	117
HMM		PGVEVTTGPIGGQIaNaVWMAIAERNLAATYNRPFGDIFDHYTYCFMGDG ++ +V+TG+LGQG++ +++++Y++++ D+++++C+MGDG	
Query	118	FV-DVATGSLGQGLG-----TACGMAYTKYLDKASYRVFCLMGDG	157
HMM		CLMEGISWEACSLAGHMqLGNWiaFYDDNcISIDGdTdIWfQEDtYakRF + +EG++WEA ++A++H++L+N++A +D NR++++G++++ + D+Y+ +	
Query	158	ESSEGSVWEAFaFASHYNLDNLVAVFDVNRLGQSGPAPLEHGADIYQNC	207
HMM		EAYGWHVIEVENdGHDvEeIcaAIEeAKaekDRPTLIICRTVIGYGSNk EA+GW++ +V DGHdVE++C A+ +A +K++PT+I ++T++G+G+PN	
Query	208	EAFGWNTYLV--DGHdVEALCQAFWQASQVKNKPTAIVAKTFKGRGIPNI	255
HMM		QGTdHWHGAPLGeD* ++ + WHG+P +++	
Query	256	EDAENWHGKVPKE 269	
HMM		*PqWePnddkIATRKASQqaleaiGPaLPefWGGsADLTSPNLtWKgmv P++++ +DKIAT K+++ AL+++G A +++ +S+D+ +S+++++ ++	
Query	311	PAYKV-GDKIATQKTYGLALAKLGRANERVIVLSGDTMNSTFSEIFRKE	358
HMM		WFMPPSISTDCynGNWwGRYIHYGIREHgmAIMNGIALHGgNFRPYGGT + + R+I++ I+E+++M++++ G+A++G+ +++++ G	
Query	359	H-----PERFIECIIAEQNMVSVLGCATRGR-TIAFAGA	392
HMM		FMMfyDYARPAIRMAALMelpVIWVWTHDSIGLGEDGPTHQPVENLAHER F+++F+++A++++RM A++ ++++++H++++ GEDG +++++E+LA+ER	
Query	393	FAAFFTRAFDQLRMGAISQANINLIGSHCGVSTGEDGVSQMALEDLAMFR	442
HMM		aIPNMwVWRPCDgNETayAWylavERehTPtLiILSRQNLQLErNPrqf +IPN +V++P+D+ T+ A YLA++++ +++++S ++ +++++ P +	
Query	443	SIPNCTVFYPSDAISTEHAIIYLAANTKGM-CFIRTSQPETAIVIY-PQEN	490
HMM		ekvaRGGYVLkDmdnePDVILIATGSEMELavaAAKlLadEGIKaRVVSM +++++++V + + + V++I++G++++A++AA+ L+ +GI +RV+++	
Query	491	FEIGQAKVVRHGvN--DKVTVIGAGVTLHEALEAADHLSQQGISVRVIDP	538
HMM		PCTeWFD.....kQDeEYReSVLPdhVPqRvAVEmGvtWCWYKYVGqq +++++D + +++++R +DH++ ++++++V ++ + + +	
Query	539	FTIKPLDAATIISAKATGGRVITVEDHYR-EGGIGEAVCAAVSREPDL	587
HMM		GaIfGMN:FGESSGRAPpevLYkMFGFTPENI* + + + + + + + + + +L+ MFG+ +I	
Query	588	VHQLAVSGVPQR---GKTSELLDMFGISTRHI 616	

DKFZphtes3\_l7n12  
-----

group: transcription factors

DKFZphtes3\_l7n12.1 encodes a novel 804 amino acid protein which is nearly identical to mouse and trout SOX-LZ.

Sox proteins belong to the HMG box superfamily of DNA-binding proteins and are involved in the regulation of developmental processes as germ layer formation, organ development and cell type specification. Deletion or mutation of Sox proteins often results in developmental defects and congenital disease in humans. Sox proteins perform their function in a complex interplay with other transcription factors in a manner highly dependent on cell type and promoter context. The new protein is related to the SOX-LZ protein and contains an additional leucin-zipper.

The new protein can find application in modulating/blocking the expression of SOX-controlled genes.

nearly identical to mouse SOX-LZ

complete cDNA, complete cds, few EST hits  
mouse and trout SOX-LZ, involved in spermatogenesis

Sequenced by GBF

Locus: unknown

Insert length: 2802 bp

Poly A stretch at pos. 2692, polyadenylation signal at pos. 2660

```
1 GGGATAGGAA AGATGAAAGG TCATGGTGAG CTTCAAGGAC ATGAAAGGTT
51 GTTGTCTCAT GTAACAATAG TAGATTGTTT TTTTTCCTAA TATTTCTAGC
101 CAGCCCCTAA GTCAGGTGAT GGAACAAATA CCTACAGTTT AGTCAGGTGA
151 AACAGGAGTG GGTGGAGGAA GGAAGAAGA AAAATGGGAA GAATGTCTTC
201 CAAGCAAGCC ACCTCTCCAT TTGCCTGTGC AGCTGATGGA GAGGATGCAA
251 TGACCCAGGA TTAAACCTCA AGGGAAAAGG AAGAGGGCAG TGATCAACAT
301 GTGGCCCTCC ATCTGCCTCT GCACCCCATTA ATGCACACAA AACCTCACAT
351 TGAGGAGCTA CCAACACTTG TCAGTACCAT TCAACAAGAT GCTGACTGGG
401 ACAGCGTTCT GTCATCTCAG CAAGAATGGG AATCAGAGAA TAATAAGTTA
451 TGTTCCTTAT ATTCCCTTCG AAATACCTCT ACCTCACCCG ATAACCCCTG
501 CGAAGGGAGT CGGGACCGTG AGATAATGAC CAGCTTACT TTTGGAAACC
551 CAGAGCGCCG CAAAGGGAGT CTGCGGATG TGCTGGACAC ACTGAAACAG
601 AAGAAGCTTG AGGAATGAC TCGGACTGAA CAAGAGGATT CCTCCTGCAT
651 GGA AAAA AACTA CTTTCAAAAG ATTTGAAGGA AAAAATGAAA AGACTAAATA
701 CAGTGAACCT TTTTCAAAAG ATTTGAAGGA CACCTGAGAG CTGGCAGAAA
751 AAGAACTGGC AGCTCTCCAC CATGATTACC CAGCTGATCA GTTTACGGGA
801 GCAGCTACTG GCAGCGCATG ATGAACAGAA AAAACTGGCA CGCTCACAAA
851 TTGAGAAACA ACGGACGCAA ATGGACCTTG CTGCGCAACA GCAAGAACAG
901 ATTCCGAGAC AACAGCAGCA ACTTCTGCAA CAGCAGCACA AAATTAATCT
951 CTGCGCAGAA CAGATCCAGG TTCAGGGTCA CATGCCTCCG CTCATGATCC
1001 CAATTTTTC ACATGACCAG CGGACTCTGG CAGCAGCTGC TGCTGCCCAA
1051 CAGGATTTC TCTTCCCCC TGGAAATACA TACAAACCAG GTGATAACTA
1101 CCCCCTACAG TTCATTCCAT CAACAATGGC AGTGCTGCT GCTTCTGGAC
1151 TCAGCCCTTT ACAGCTCCAG CAGCTCATG CCGCTCAGCT GGCCAGCATG
1201 CAGGTGTCTC CTGGAGCAAA GATGCCATCA ACTCCACAGC CACCAAAACAC
1251 AGCAGGGACG GTCTCACCTA CTGGGATAAA AAATGAAAAG AGAGGGACCA
1301 GCCCTGTAAC TCAAGTTAAG GATGAAGCAG CAGCACAGCC TCTGAATCTC
1351 TCATCCCGAC CCAAGACAGC AGAGCCTGTA AAGTCCCAAA CGTCTCCAC
1401 CCAGAACCTC TTCCAGCCA GCAAAACCAG CCCTGTCAAT CTGCCAAACA
1451 AAAGCAGCAT CCCTAGCCCC ATTGGAGGAA GCCTGGGAAG AGGATCCTCT
1501 TTAGGTAAAT GGAAGAAGTCA ACACAGGAA GAGACTTACG AATTAGATAT
1551 CCTATCTAGT CTCAACTCCC CTGCCCTTTT TGGGGATCAG GATACAGTGA
1601 TGAAGCCCAT TCAGGAGGCG CGGAAGATGC GAGAGCAGAT CCAGCGGGAG
1651 CAACAGCAGC AACAGCCACA TGGTGTGAC GGGAACTGT CCTCCATAAA
1701 TAATATGGGG CTGAACAGCT GCAGGAATGA AAAGGAAAGA ACGCGCTTGT
1751 AGAATTTGGG GCCCGACTTA ACGGGAAAGT CAAATGAAGA TGGAAAAGTG
1801 GGCCAGGTG TCATCGACCT TACTCGGCCA GAAGATGCAG AGGGAAGTAA
1851 AGCAATGAAT GGCTCTGGAG CTAACTACA GCAGTATTAT TGTGGGCCAA
1901 CAGGAGGTGC CACTGTGGCT GAAGCAGGAG TCTACAGGGA CGCCCGCGGC
1951 CGTGGCCAGCA GCGAGCCACA CATTAGCGGA CCAATGAATG CATTCTGGT
2001 TTGGGCAAAAG GATGAGAGGA GAAAAATCCT TCAGGCCTTC CCCGACATGC
2051 ATAACCTCAA CATTAGCAAA ATCTTAGGAT CTGCTGGAA ATCAATGTCC
2101 AACCAGGAGA AGCAACCTTA TTATGAAGAG CAGGCCCGGC TAAGCAAGAT
2151 CCACTTAGAG AAGTACCCAA ACTATAAATA CAACCCCGCA CCGAAACGCA
2201 CCGCATTGTG TGATGGCAAA AAGCTTCGGA TTGGGAGTA TAAGCAACTG
2251 ATGAGGTCTC GGAGACAGGA GATGAGGAG TTCTTACTG TGGGGCAACA
2301 GCCTCAGATT CCAATCACA CAGGAACAGG TGTGTGTAT CCTGGTGCTA
2351 TCATATGGC AACTACCACA CCATCGCCTC AGATGACATC TGACTGTCTT
```

```

2401 AGCACCTCGG CCAGCCCGGA GCCCAGCCTC CCGGTCATCC AGAGCACTTA
2451 TGGTATGAAG ACAGATGGCG GAAGCCTAGC TGGAAATGAA ATGATCAATG
2501 GAGAGGATGA AATGGAATG TATGATGACT ATGAAGATGA CCCCAATCA
2551 GACTATAGCA GTGAAATGA AGCCCGGAG GCTGTCAGT CCAACTGAGG
2601 AGTTTTTGTG TGCTGAATTA AAGTACTCTG ACATTTCACC CCCCTCCCA
2651 ACAAGAGATT ATTAAGAGC CCGCATGCAT TTGTGGCTCC ACAATTAAAA
2701 AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA
2751 AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA
2801 AA

```

## BLAST Results

No BLAST result

## Medline entries

95311974:  
A gene that is related to SRY and is expressed in the testes encodes a leucine zipper-containing protein.

96032826:  
The Sry-related HMG box-containing gene Sox6 is expressed in the adult testis and developing nervous system of the mouse.

## Peptide information for frame 1

ORF from 184 bp to 2595 bp; peptide length: 804  
Category: strong similarity to known protein

```

1 MGRMSSKQAT SPFACAAADGE DAMTQDLTSR EKEEGSDQHV ASHLPLHPIM
51 HNKPHSEELP TLVSTIQQDA DWDSVLSSQQ RMESNNKLC SLYSFRNTST
101 SPHKPDEGSR DREIMTSVTF GTPERRKGS L ADVVDTLKQ KLEEMTRTEQ
151 EDSSCMEKLL SKDWKEKMER LNTSELLGEI KGTPESLAEK ERQLSTMITQ
201 LLSLREQLLA AHDEQKKLAA SQIEKQRQQM DLARQQEQI ARQQQLLQQ
251 QHKINLLQQQ IQVQGHMPP L MIPFPHDQR TLAAAAAQ GFLFPFGITY
301 KPGDNYPVQF IPSTMAAAAA SGLSPLQLQQ LYAAQLASMQ VSPGAKMPST
351 PQPPNTAGTV SPTGIKNEKR GTSPTVQVKD EAAQPLNLS SRPKTAEPVK
401 SPTSPTQNL F PASKTSPVNL PNKSSIPSPI GGLGRGSSL GKWKSHQEE
451 TYELDILSSL NSPALFGQDQ TVMKAIQEAR KMREIQREQ QQQPHGVG
501 KLSSINNMGL NSCRNEKERT RFENLGPQLT GKSNEGKLG PGVIDLTRPE
551 DAEGSKAMNG SAAKLQQYIC WPTGGATVAE ARVYRDARGR ASSEPHIKRP
601 MNAFMVWAKD ERRKILQAFP DMHNSNISKI LGSRMKSMN QEKQPYVEQ
651 ARLSKIHLEK YPNYKIKPRP KATCIVDGKK LRIGETKQLM RSRQEMROF
701 FTVGQQPQIP ITTGTGVVYP GAITMATTTT SPQMTSDCSS TSASPEPSLP
751 VIQSTYGMKT DGGSLAGNEM INGEDEMERY DDYEDDFKSD YSSENEAPEA
801 VSN

```

## BLASTP hits

Entry MMSOXL22.1 from database TREMBL:  
product: "SOX-L2"; Mouse mRNA for SOX-L2, complete cds.  
Score = 3910, P = 0.0e+00, identities = 764/801, positives = 774/801

Entry I51083 from database PIR:  
SOX-L2 - rainbow trout  
Score = 1774, P = 1.1e-287, identities = 365/532, positives = 431/532

Entry S59121 from database PIR:  
SOX6 protein - mouse  
Score = 2319, P = 1.2e-240, identities = 489/660, positives = 527/660

Entry AB006330.1 from database TREMBL:  
gene: "mSox5L"; product: "SOX5"; Mus musculus mSox5L mRNA, complete cds.  
Score = 1212, P = 8.9e-209, identities = 274/457, positives = 324/457

Entry MMU010604.1 from database TREMBL:  
gene: "sox5"; product: "L-Sox5 protein"; Mus musculus mRNA for transcription factor L-Sox5  
Score = 879, P = 4.2e-195, identities = 190/281, positives = 218/281



Alert BLASTP hits for DKFZphtes3\_17n12, frame 1

No Alert BLASTP hits found

Pedant information for DKFZphtes3\_17n12, frame 1

Report for DKFZphtes3\_17n12.1

[LENGTH] 804  
[MW] 89332.69  
[pI] 6.97  
[HOMOL] TREMBL:MMSOXLZ2\_1 product: "SOX-LZ"; Mouse mRNA for SOX-LZ, complete cds. 0.0  
[FUNCAT] 04.05.01.04 transcriptional control [S. cerevisiae, YKL032c] 8e-07  
[FUNCAT] 30.10 nuclear organization [S. cerevisiae, YKL032c] 8e-07  
[FUNCAT] 01.07.07 regulation of vitamins, cofactors, and prosthetic groups [S. cerevisiae, YPR065w] 5e-06  
[FUNCAT] 03.04 budding, cell polarity and filament formation [S. cerevisiae, YBR089c-a] 7e-06  
[FUNCAT] 30.13 organization of chromosome structure [S. cerevisiae, YBR089c-a] 7e-06  
[FUNCAT] 03.01 cell growth [S. cerevisiae, YBR089c-a] 7e-06  
[FUNCAT] 03.16 dna synthesis and replication [S. cerevisiae, YMR072w] 2e-04  
[FUNCAT] 30.16 mitochondrial organization [S. cerevisiae, YMR072w] 2e-04  
[SCOP] dlhmf\_1.20.1.1.1 HMGI, fragments A and B [rat/hamster (Rattus 1e-13  
[SCOP] dllef\_1.20.1.1.6 Lymphoid enhancer-binding factor, LEF1 [mous 4e-15  
[SCOP] dlhry\_1.20.1.1.4 SRY [Human (Homo sapiens) 7e-17  
[PIRKB] DNA binding 4e-94  
[PIRKB] T-cell receptor 4e-07  
[PIRKB] leucine zipper 1e-38  
[PIRKB] alternative splicing 2e-07  
[PIRKB] transcription factor 4e-16  
[PIRKB] transcription regulation 1e-12  
[SUPFAM] HMG box homology 0.0  
[SUPFAM] unassigned HMG box proteins 4e-94  
[PROSITE] ATP\_GTP\_A 1  
[PROSITE] LEUCINE\_ZIPPER 1  
[PROSITE] MYRISTYL 6  
[PROSITE] AMIDATION 1  
[PROSITE] CAMP\_PHOSPHO\_SITE 2  
[PROSITE] CK2\_PHOSPHO\_SITE 14  
[PROSITE] PKC\_PHOSPHO\_SITE 10  
[PROSITE] ASN\_GLYCOSYLATION 6  
[PFAM] HMG (high mobility group) box  
[KW] Irregular  
[KW] 3D  
[KW] LOW\_COMPLEXITY 13.81 %  
[KW] COILED\_COIL 3.48 %

SEQ MGRMSSKQATSPFACAAAGDAMTQDLTSREKEEGSDQHVASHLPHPIMHNKPHSEELP  
SEG .....  
COILS .....  
lnhm- .....  
SEQ TLVSTIQQADWDVSLSSQRMESNNKLCSLYSFRNTSTSPHKPDGSRDREIMTSVTF  
SEG .....  
COILS .....  
lnhm- .....  
SEQ GTPERRKGLADVDTLKQKKLEEMTRTEQEDSSCKEKLKSKDWKEKMERLNTSELLGEI  
SEG .....  
COILS .....  
lnhm- .....  
SEQ KGTPESLAEKERQLSTMITQLISLREQLAAHDEQKKLAASQIEKQRMQMDLARQQQEQEI  
SEG .....  
COILS .....  
lnhm- .....  
SEQ ARQQQQLLQQQHKNLLQQQIQVQGHMPLMIPFPHDQRTLAAAAAQQGFLFPPGITY  
SEG .....  
COILS .....  
lnhm- .....  
SEQ KPGDNYPVQFIPSTMAAAAAGLSPLQLQQLYAAQLASMQVSPGAKMPSTPQPPNTAGTV  
SEG .....  
COILS .....  
lnhm- .....

```

COILS .....
1nhm- .....

SEQ      SPTGIKNEKRGTSPTVQVKDEAAQPLNLSSRPKTAEPVKSPTSPTONLFPASKTSPVNL
SEG      .....
COILS    .....
1nhm-    .....

SEQ      PNKSSIPSPIGGSLGRGSSLGKWKSOHQEETVELDILSSLNSPALFGDQDTVMKAIQEAR
SEG      .....
COILS    .....
1nhm-    .....

SEQ      KMREQIQREQQQQPHGVDGKLSSINNMGNSCRNEKERTRFENLGPQLTGKSNEDGKLG
SEG      .....
COILS    .....
1nhm-    .....

SEQ      PGVIDLTRPEDAEGSKAMNGSAAKLQYYCWP TGGATVAEARVYRDARGRASSEPHIKRP
SEG      .....
COILS    .....
1nhm-    .....CCC

SEQ      MNAFMVWAKDERRKILQAFPMHNSNISKILGSRWKSMSNQEKPYYEEQARLSKIHLEK
SEG      .....
COILS    .....
1nhm-    .....CCCHHHHHHHHHHHHHHTTTTCCHHHHHHHHHHHTTTTTHHHHHHHHHHHHHHHHHHH

SEQ      YPNYKYKPRPKRTCIVDGKKLRIGEYKQLMRSRRQEMRQFETVGGQPQIPITTTGTGVVYP
SEG      .....
COILS    .....
1nhm-    .....HHHTTTTTTT

SEQ      GAITMATTTPSPQMTSDCSSTSASPEPSLPVIQSTYGMKTDGGS LAGNEMINGEDEMEMY
SEG      .....
COILS    .....
1nhm-    .....

SEQ      DDYEDDPKSDYSSENEAPAVSAN
SEG      .....
COILS    .....
1nhm-    .....

```

Prosites for DKF2phtes3\_17n12.1

PS00001	97->101	ASN_GLYCOSYLATION	PDOC00001
PS00001	172->176	ASN_GLYCOSYLATION	PDOC00001
PS00001	388->392	ASN_GLYCOSYLATION	PDOC00001
PS00001	422->426	ASN_GLYCOSYLATION	PDOC00001
PS00001	559->563	ASN_GLYCOSYLATION	PDOC00001
PS00001	626->630	ASN_GLYCOSYLATION	PDOC00001
PS00004	126->130	CAMP_PHOSPHO_SITE	PDOC00004
PS00004	369->373	CAMP_PHOSPHO_SITE	PDOC00004
PS00005	5->8	PKC_PHOSPHO_SITE	PDOC00005
PS00005	28->31	PKC_PHOSPHO_SITE	PDOC00005
PS00005	94->97	PKC_PHOSPHO_SITE	PDOC00005
PS00005	136->139	PKC_PHOSPHO_SITE	PDOC00005
PS00005	203->206	PKC_PHOSPHO_SITE	PDOC00005
PS00005	299->302	PKC_PHOSPHO_SITE	PDOC00005
PS00005	390->393	PKC_PHOSPHO_SITE	PDOC00005
PS00005	512->515	PKC_PHOSPHO_SITE	PDOC00005
PS00005	530->533	PKC_PHOSPHO_SITE	PDOC00005
PS00005	692->695	PKC_PHOSPHO_SITE	PDOC00005
PS00006	28->32	CK2_PHOSPHO_SITE	PDOC00006
PS00006	129->133	CK2_PHOSPHO_SITE	PDOC00006
PS00006	146->150	CK2_PHOSPHO_SITE	PDOC00006
PS00006	148->152	CK2_PHOSPHO_SITE	PDOC00006
PS00006	154->158	CK2_PHOSPHO_SITE	PDOC00006
PS00006	186->190	CK2_PHOSPHO_SITE	PDOC00006
PS00006	203->207	CK2_PHOSPHO_SITE	PDOC00006
PS00006	221->225	CK2_PHOSPHO_SITE	PDOC00006
PS00006	520->524	CK2_PHOSPHO_SITE	PDOC00006
PS00006	533->537	CK2_PHOSPHO_SITE	PDOC00006
PS00006	547->551	CK2_PHOSPHO_SITE	PDOC00006
PS00006	577->581	CK2_PHOSPHO_SITE	PDOC00006
PS00006	639->643	CK2_PHOSPHO_SITE	PDOC00006
PS00006	793->797	CK2_PHOSPHO_SITE	PDOC00006
PS00008	182->188	MYRISTYL	PDOC00008
PS00008	431->437	MYRISTYL	PDOC00008

WO 01/12659

PCT/IB00/01496

PS00008	437->443	MYRISTYL	PDOC00008
PS00008	509->515	MYRISTYL	PDOC00008
PS00008	575->581	MYRISTYL	PDOC00008
PS00008	762->768	MYRISTYL	PDOC00008
PS00009	677->681	AMIDATION	PDOC00009
PS00017	526->534	ATP_GTP_A	PDOC00017
PS00029	187->209	LEUCINE_ZIPPER	PDOC00029

Pfam for DKF2phtes3\_17n12.1

HMM_NAME	HMG (high mobility group) box		
HMM	*PKRPMNAYMLWQEMRekIKaENPNdMhNtEISKMiGEMWKnMsEEEkM +KRPMNA+M+W+++ R+KI + P DMHN++ISK++G +WK+MS +EK+		
Query	597	IKRPMNAFMVWAKDERRKILQAFP-DMHNSNISKILGSRWKSMSNQEKQ	644
HMM	PYEdMAeeEKqRYMKEMPeYK*		
Query	645	PYYEEQARLSKIHLEKYPNYK	665

DKFZphtes3\_17n18  
-----

group: intracellular transport and trafficking

DKFZphtes3\_17n18 encodes a novel 782 amino acid protein with weak partial similarity to known proteins.

The novel protein contains a ATP/GTP-binding site motif A (P-loop) and a TonB-dependent receptor protein signature 1. In *E. coli*, the tonB protein interacts with outer membrane receptor proteins that mediate uptake of specific substrates into the periplasmic space. In the absence of tonB these receptors bind their substrates but do not carry out active transport. The novel protein seems to be involved in ATP-dependent transport of substances into the cell.

The new protein can find application in modulation of cell-permeability and transport of suitable substrates into the cell.

unknown receptor

protein contains TONB\_DEPENDENT\_REC\_1 Pattern and ATP\_GTP\_A Pattern,

Sequenced by GBF

Locus: unknown

Insert length: 2853 bp

Poly A stretch at pos. 2806, no polyadenylation signal found

```
1  GTCCTTTTAA  GTCAGTAAAT  TGAAC TAAGT  CGGTTATTCG  GCAAGCAGTT
51 CCTATAAAAA  ACTACATGGC  TAAGTTCTT  AATGATTGAC  CACAAGCAGA
101 TCTTTCACCC  TCGGATCTCT  AGCTACAAAA  GGTCCCCACA  CTGAAGAAGC
151 CACTACCTCC  ACCACCACCA  GCACCACCA  GTCCAGTGCT  GCTGGCAACC
201 ACTGGGGCAG  CCAAGCGCTC  CACCCTCTCT  CCCACCATGG  CCGTCAGGT
251 GCGCACCAC  CAGGAGACCC  TGAACAGGTT  TCAGCAGCAG  TCCATCCACC
301 TGCTGACGGA  GCTCCTCAGA  CTGAAGATGA  AGGCCATGGT  GGAGTCTATG
351 TCGGTGGGTG  CCAACCCCTT  GGACATCACC  AGGCGCTTTG  TGGAGGCCAG
401 CCAGCTCCTC  CACCTCAATG  CCAAGGAGAT  GGCCTTCAAC  TGCCTGATCA
451 GCACAGCCGG  GAGAAGTGCC  TACAGCAGCG  GACAGTTGTG  GAAAGAGTCC
501 CTCGCAAAAC  TGTCCGCCAT  TGGGGTGAAC  TCGCCTTACC  AGCTGATCTA
551 CCACCTCTCC  ACAGCCTGTC  TGAGCTTTTC  TCTCTCTGCT  GGAAAAAGAG
601 CCAAGAAGAA  AATAGGCAAA  TCTAGAACTA  CAGAAGATGT  CAGCATGCCG
651 CCCCTGCATC  GAGGAGTGGG  AACCCCTGCC  AACAGCCTGG  AGTTCAGCGA
701 CCCCTGCCCT  GAGGCCCGGG  AGAAGCTGCA  GGAGTTGTGT  CGCCACATAG
751 AAGCTGAAAG  GGCCACATGG  AAAGGGAGGA  ATATCTCCTA  CCCCATGATC
801 TTACGAAACT  ACAAGGCAAA  GATGCCCTCT  CATCTAATGT  TGGCCCCGAA
851 AGGAGACTCT  CAGACCCCGG  GTTTACATTA  CCTCCCACT  GCAGTGCTCT
901 AGACTCTCAG  CCCCACTCT  CACCCATCTT  CTGCCAACCA  TCATTTCAGT
951 CAGCATTGTC  AAGAGGGGAA  GGCACCCAA  AAGCCCTTCA  AGTTTCATTA
1001 CACCTTCTAT  GATGGCTCCT  CCTTCGTTA  CTATCCCTCT  GGAACGCTCG
1051 CTGTATGTCA  GATCCCCACA  TGCTGCAGAG  GGAGAACCAT  CACCTGCCTC
1101 TTTAATGACA  TACCTGGATT  CTCCTTGCTG  GCCCTATTCA  ATACTGAAGG
1151 CCAGGGCTGT  GTTCACTACA  ACCTAAAAAC  CAGTTGCCCA  TATGCTTAA
1201 TCTTGGATGA  GGAAGTGGG  ACCACCAATG  ACCAGCAGGG  CTATGTAGTC
1251 CACAAGTGGG  GCTGGAATTC  CAGGACAGAG  ACCCTGCTTT  CCTGGAATA
1301 CAAGGTGAAT  GAGGAAATGA  AACTAAAGGT  ACTGGACAG  GACTCCATCA
1351 CAGTCACTTT  CACCTCCCTG  AATGAGACAG  TAACACTCAC  TGTGTCGGCC
1401 AACATTTGTC  CCCATGGAAT  GGCATATGAC  AACCGGCTGA  ACCGCAGAAT
1451 CAGCAACATG  GACGACAAGG  TGTATAAGAT  GAGCCGAGCC  CTGGCTGAGA
1501 TCAAGAGCGG  GTTTCAGAAG  ACAGTGACTC  AGTTCAATTA  TTCTATCTTG
1551 CTGGCCGCAG  CTCTGTTTAC  CATTGAATAT  CCCACCAAAA  AGGAGGAGGA
1601 AGAATTTGTT  CGGTTCAGAA  TGAGATCCAG  AACTCATCCC  GAGCGGCTCC
1651 CCAAGCTAAG  TTTATACTCA  GGAGAAAGTC  TTTTACGATC  TCAGTCAGGC
1701 CACCTGGAAT  CCTCAATTGC  AGAGACTTTG  AAGGATGAGC  CTGAGTCTGC
1751 TCCTGTGAGC  CCAGTTCGGA  AGACCACCAA  AATCCACACC  AAAGCCAAAG
1801 TCACATCCAG  AGGGAAGGCC  CGCGAGGGCC  GCAGCCCCAC  CAGGTGGCGG
1851 GCCTTGCCCT  CAGACTGCCC  GCTGGTGCTG  CGGAAGCTCA  TGCTCAAGGA
1901 AGACACCCGT  GCTGGTGCA  AGTGCTGGT  GAAGGCCGCC  CTGGTCTCTG
1951 ACGTGGAGCT  GGAGCGTTTC  CTGTTGGCGC  CCGGAGACCC  CAGCCCAAGT
2001 CTGGTGTGTT  GGATCATCTC  AAGCCAGAAC  TACACCAGCA  CTGGGCAGCT
2051 CCAGTGGCTG  CTGAACACTC  TCTACAACA  CCAGCAGCGG  GGCCGTGGCT
2101 CCCCTGTCAT  CCAGTGCCTG  TATGACTCT  ACCGCTGCTG  GCAGTATGAC
2151 CTGGACAGCC  CCCTGCAGGA  GGACCTCTCC  CTGATGGTGA  AGAAGAATCT
2201 TGTGGTGCG  GGGATGATTC  TGATGTTTC  CGGGGGGGA  CTCATTTTTC
2251 GGGGCCGTGT  TTTGAATGGA  TATGGCTTCA  GCAAGCAGAA  TCTGCTGAAA
2301 CAGATCTTCC  GGTCTCAACA  GGATTACAAG  ATGGGCTACT  TCCTGCCGGA
2351 TGACTACAAA  TTCAGTGTTT  CCAACTCTGT  CCTGAGCCTG  GAGGATTCTG
2401 AATCAGTCAA  GAAAGCCGAG  TCAGAAGATA  TCCAAGGAAG  CAGCTCCTCA
```

WO 01/12659

PCT/IB00/01496

```
2451 TTGGCCCTGG AAGACTATGT GGAGAAGGAG TTATCTCTGG AGGCTGAGAA
2501 GACAAGAGAG CCTGAAGTGG AGCTACATCC TCTCAGCAGG GACAGCAAGA
2551 TAACTAGTTG GAAGAAGCAG GCCTCAAGA AGTAGCGCCA TCCTGGCAGC
2601 AGCCAAGTGA GCCAGGCCCC GGGCCGGGGT GCTGGGGCTT CTGCCCAGCC
2651 CAGCCCTGCC TCCCGGGTCT CCCACCCTGT CCTCCAAGCT TCTATAATAA
2701 ACCAGCGGGC CTCCAGCATT GGGGTGAGCG TCTGGGGAAG GACAAAAAAA
2751 AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAGGGG
2801 CGGCGGAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAGGGCGG
2851 CCG
```

BLAST Results

No BLAST result

Medline entries

No Medline entry

Peptide information for frame 3

ORF from 237 bp to 2582 bp; peptide length: 782  
Category: putative protein  
Prosite motifs: ATP\_GTP\_A (122-130)  
TONB\_DEPENDENT\_REC\_1 (1-44)

```
1 MARQVRTHQE TLNRFOQOSI HLLTELLRLK MKAMVESMSV GANPLDITRR
51 FVEASQLLHL NAKEMAFNCL ISTAGRSGYS SGQLWKESLA NMSAIGVNSP
101 YQLIYHSSTA CLSFSLSAGK EAKKIKGKSR TTEDVSMPPPL HRVGTPANS
151 LEFSDPCPEA REKLQELCRH IEAERATWKG RNISYPMILR NYKAKMPSHL
201 MLARKGDSQT PGLHYPTTAG AQTLSPTSHS SSANHHFSQH COEGKAPKKA
251 FKPHYTFYDG SSFVYYPGSG VAVCQIPTCC RGRITITCLFN DIPGFSLLAL
301 FNTEGQGCYH YNLKTSCPYV LILDEEGGTT NDQQGYVVHK WSWTSRTETL
351 LSLEYKVNEE MKLVVLGQDS ITVTFTSLNE TVTLTVSANN CPHGMAYDKR
401 LNRISNMDD KVKMSRALA EIKKRFQKTV TQFINSILLA AGLFTIEYPT
451 KKEEEFVRP KMRSRTHPER LPKLSLYSGE SLLRSQSGHL ESSIAETLKD
501 EPESAPVSPV RKTTHHTKA KVTSRGKARE GRSPTRWAAL PSDCPLVLAK
551 LMLKEDTRAG CKCLVKAPLV SDVELERFLL APRDPQVQLV FGISSQNYT
601 STGQLQWLLN TLYNHQQRGR GSPCIQCRDY SYRLQYDLD SPLQEDPPLM
651 VKKNSVVQGM ILMFAGGKLI FGGRVLNGYG LSRQNLKQI FRSQDDYKMG
701 YFLPDDYKFS VPHSVLSLED SESVKRAESE DIQSSSSSLA LEDYVEKELS
751 LEAENTREPE VELHPLSRDS KITSWKKQAS KK
```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphtes3\_17n18, frame 3

No Alert BLASTP hits found

Pendant information for DKFZphtes3\_17n18, frame 3

Report for DKFZphtes3\_17n18.3

```
{LENGTH} 782
{MW} 88030.16
{pI} 9.22
{BLOCKS} BL00286 Squash family of serine protease inhibitors proteins
{PROSITE} ATP_GTP_A 1
{PROSITE} MYRISTYL 4
{PROSITE} CAMP_PHOSPHO_SITE 3
{PROSITE} CK2_PHOSPHO_SITE 14
{PROSITE} PROKAR_LIPOPROTEIN 1
{PROSITE} TONB_DEPENDENT_REC_1 1
{PROSITE} PKC_PHOSPHO_SITE 10
{PROSITE} ASN_GLYCOSYLATION 4
{KW} Alpha_Beta
```

SEQ	MARQVTRHQTETLNRFFQOOSILHLTELRLMLKMKAMVESMSVGPANPLOITRRFVEASQLLHL
PRD	ccccchhh
SEQ	NKEMAFNCLISITAGRSVSGYLKVESGLKVESLAMSIAIVGNVSPQLIYHSSTACLFSLSLACK
PRD	hhhhhhhhhhhhhhhhccccccccccccchhhhhhhhhhhhhhhcccccccccccccccccccccccc
SEQ	EAKKIKGSRITTEDVSMPLLRHGVGTFPANSLFSDPCPEAREKLQELCRHIEAERATWKG
PRD	hh
SEQ	RNISYPMILRNKYAKMPSHMLMARKGDSQTPGLHYFPPTAGQATLSPTSHSSANHHFSQH
PRD	ccccchhhhhhhhhhhhhhhcc
SEQ	CQEGKAPKKAFFHYTIFYDGSSFVYYPSGNVAVQCITPCRCGRITICLFDNIPGFSLLAL
PRD	ccccchhh
SEQ	FNTEGGQGVHYNLKTSCTPYVILDEEGTTCNDQQQVYVHKWSWTSRTETLLSLEYKVNEE
PRD	ccchhhhhhhhhhhhhhh
SEQ	MKLVLHGQDSITVTFETSLNETVTLTYSANNCPHGMAYDKRLNRRISNMDDKVKYMSRALA
PRD	hh
SEQ	EIKKRFQKTVTQFINSILLAAGLFTI EYPTKKEEEFVRFKMRSRTHPERLPKLSLYSGE
PRD	hh
SEQ	SLLRSSQGHLESSIAHETLKDEPESAPVPRKTTKIHTKAKVTSRGKAREGRSPTRWAAL
PRD	ccccccccccccchhh
SEQ	PSDCPLVLRKLMLEKEDTRAGCKCLVRAPLVDSVLELRFHLLHCCPPCSQVLFVFGIISQNTV
PRD	ccccchhh
SEQ	STGOLQWLNTLYNHQGRGSCPCIQRYDSYRLLYQDLSFLQEDPPLVMKKNVSNVQGM
PRD	chhh
SEQ	ILMFAGGKCLIFGGRVLNGLYSKQNLKLQIFRSQQOYKMGYFLPDDYFVSVPNSVLSLED
PRD	heccccccccccccccccccccchhh
SEQ	SESVKAEESDIOGSSSIALEDYVEKELSLEAKTREPVEVHLPLSRDSKITSWKKQAS
PRD	chhh
SEQ	KK
PRD	cc

Prosites for DKFZphtes3\_17n18.3

PS000001	91->95	ASN_GLYCOSYLATION	PDOC000001
PS000001	182->186	ASN_GLYCOSYLATION	PDOC000001
PS000001	379->383	ASN_GLYCOSYLATION	PDOC000001
PS000001	598->602	ASN_GLYCOSYLATION	PDOC000001
PS000004	403->407	CAMP_PHOSPHO_SITE	PDOC000004
PS000004	511->515	CAMP_PHOSPHO_SITE	PDOC000004
PS000004	652->656	CAMP_PHOSPHO_SITE	PDOC000004
PS000005	48->51	KPC_PHOSPHO_SITE	PDOC000005
PS000005	177->180	KPC_PHOSPHO_SITE	PDOC000005
PS000005	344->347	KPC_PHOSPHO_SITE	PDOC000005
PS000005	450->453	KPC_PHOSPHO_SITE	PDOC000005
PS000005	497->500	KPC_PHOSPHO_SITE	PDOC000005
PS000005	513->516	KPC_PHOSPHO_SITE	PDOC000005
PS000005	523->526	KPC_PHOSPHO_SITE	PDOC000005
PS000005	631->634	KPC_PHOSPHO_SITE	PDOC000005
PS000005	723->726	KPC_PHOSPHO_SITE	PDOC000005
PS000005	774->777	KPC_PHOSPHO_SITE	PDOC000005
PS000006	7->11	CK2_PHOSPHO_SITE	PDOC000006
PS000006	131->135	CK2_PHOSPHO_SITE	PDOC000006
PS000006	256->260	CK2_PHOSPHO_SITE	PDOC000006
PS000006	329->333	CK2_PHOSPHO_SITE	PDOC000006
PS000006	345->349	CK2_PHOSPHO_SITE	PDOC000006
PS000006	377->381	CK2_PHOSPHO_SITE	PDOC000006
PS000006	406->410	CK2_PHOSPHO_SITE	PDOC000006
PS000006	450->454	CK2_PHOSPHO_SITE	PDOC000006
PS000006	466->470	CK2_PHOSPHO_SITE	PDOC000006
PS000006	493->497	CK2_PHOSPHO_SITE	PDOC000006
PS000006	497->501	CK2_PHOSPHO_SITE	PDOC000006
PS000006	571->575	CK2_PHOSPHO_SITE	PDOC000006
PS000006	693->697	CK2_PHOSPHO_SITE	PDOC000006
PS000006	717->721	CK2_PHOSPHO_SITE	PDOC000006
PS000008	145->151	MYRISTYL	PDOC000008
PS000008	327->333	MYRISTYL	PDOC000008
PS000008	592->598	MYRISTYL	PDOC000008
PS000008	734->740	MYRISTYL	PDOC000008

WO 01/12659

PCT/IB00/01496

PS00013	101->112	PROKAR_LIPOPROTEIN	PDOC00013
PS00017	122->130	ATP_GTP_A	PDOC00017
PS00430	1->44	TONB_DEPENDENT_REC_1	PDOC00354

(No Pfam data available for DKF2phtes3\_17n18.3)

DKFZphtes3\_18f3  
-----

group: testes derived

DKFZphtes3\_18f3 encodes a novel 248 amino acid protein with partial similarity to human TNF-inducible protein CG12-1.

The novel protein contains two leucine zippers.

No informative BLAST results: No predictive prosite, pfam or SCOP motive.

The new protein can find application in studying the expression profile of testis-specific genes.

similarity to TNF-inducible protein CG12-1

Sequenced by MediGenomix

Locus: unknown

Insert length: 4608 bp

Poly A stretch at pos. 4570, polyadenylation signal at pos. 4550

```
1 GACAGAAGTG AATGGGAATG GAGAGGCCGG CGGCCCGGGA GCCGCATGGG
51 CCCGACGCCG TCGCGCCGCTT CCAGGGACTG CTGCTGGACC GCCGAGGCCG
101 GCTGCACCGC CAGGTGCTGC GCCTGCGCGA GGTGGCCCGG CGCCTGGAGC
151 GCCTGCGCAG GCGCTCCCTC GTAGCCAAAG TGGCCGCCAG CTCCTGAGC
201 GCAACGGGCG CCCTCGCCGC CATCGTGGGG CTCTCGCTCA GCCCGGTAC
251 CCTGGGGACC TCGCTGCTGG TGTGGCCGCT GGGGCTGGGG GTGGCCACAG
301 CCGGAGGGGC CGTCACCATC ACGTCCGATC TCTCGCTGAT CTTCTGCAAC
351 TCCCGGGAGC TCGGGAGGGT GCAGGAGATC GCGGCCACCT GCCAGGACCA
401 GATGCGAGAG ATCCTGAGCT GCCTCGAGTT TTTCTGCCGC TGGCAGGGCT
451 GCGGGGACCG CCAGCTGCTG CAGTGGCGGA GGAACGCCCT CATCGCCCTG
501 TACAATTCTG TCTACTTCAT CGTCTTCTTT GGCTCACGTG GCTTCTCAT
551 CCCAGGGCGG GCGGAGGGGG ACACCAAGGT TAGCCAGGCC GTGCTGAAGG
601 CCAAGATTCA GAAACTGGCC GAGAGCCTGG AGTCTGACAC CGGGGCTCTG
651 GACGAATCA GCGAGCAGCT GGAGTCTCGG GTTCAGCTCT GCACCAAGTC
701 CAGTCGTGGC CACGACCTCA AGATCTCTCG TGACACGCGT GCAGGGCTGT
751 TTTTCTGAGA ACATCCTTTC CCCCTAATGA CCGAGGCCAG CAATCATCC
801 TCATGGGATG CTCAGAAATT TGTAGCTCCC TTAGGAAAAC ACCAAGCTGG
851 GTTAGGAGCC GAAGGCCAAG GATGAGAAAA ACTGTTTTTG AAGTGGGCG
901 GTCCCAAGAG CCCTTCTTTT CCCATCACTG TGACATCTGC CTGGGCTTGA
951 GTGCTACGGA CTTTTCAGTC TTCCTAGTGG AAAAATGTGA CCCAAAACCT
1001 CCTTTTCTTT TATCAAAAAC TTCTGTGCTA AACACAGCTG GGCAGGCCCT
1051 CCTGTTTTAA AGTTATTTTC GGGTCCCTGA CCCTGCCCTG GTGGCTTGGC
1101 CTGAGACTGG AGAGAGTGCC ATCCTCTGGG TCCTCTCCAA GTCTACTAG
1151 TCTTTGAAAT CCTCAAAATG TGGGTGAGGA AGCCATTTCG CTCTATTCCA
1201 GAATTTCTGA TACAAGAAGC TCCAGAAATC AGAGCAAAAT AGCCCTTCTC
1251 TGAACGTTGT AGGATGGTTC AGAACCCAGA GAGGACCCCT GTGCTGATAT
1301 CTCCTCCTCT TCCCTTTCCC CTCAGCTTAC TTACTCCAGT ATGCGGCCCT
1351 GGTATGAAGT AGGCCCTTCC TCACTGGCTC CCAATCCAGT CCTCAAAGTA
1401 CTCAGAGGGG AAGCCCGTGA AGCCGTCATC TAACTCCTCG TCCTCTACAT
1451 GAAGCTGAGG CCCAGATAGA TGGAGGCACT GCCAACTTCA TTCCCGACA
1501 TCATTCTGCT CAGAAGAGAG TGATGGGTTT TGAGTTAGAC AGTCTGGGCG
1551 TTGAGACAGG CTTTGTCACT ACTGTGTGAG TGTAGCCACC TAATCTCTCT
1601 GAGACTGTGT AAAACAAGA TGATAAAATC TCACCTGTTT GTGAGATATT
1651 AATGAGCCA AAGTGCCTAG CATGATGGTG CTGGCTCATA TAGTGTAGTC
1701 CTGGAATGG CAAATTAACA TCACCCAGGA ACTTGTTAGA AAGGCAAAAT
1751 CTTGACACA ACCCTCCTGA TTTATGGAAT CAGAACTCT GGCTGTGGGG
1801 CCCAGCAACC TGAGTTTAAA CAATTCTCTT GGTGGTTCT GCGGCACACT
1851 AAGTTTGAA AATCACTACA ACAAAATGCTA ACTTCTAATC CCCTTGATGA
1901 CTTTCAACA AGTCTCAGG CTTCTCTAGG GACTCCATGG TCTTCAGAGT
1951 CGTTACAGA TGACCAAGGA CAGACTGTGT CCCAGAGCC AAAATGAGAG
2001 AGAGAGAGAG AGCAGCGGTA CGTGACCCCT GGGGCACTGT CTCACCGTAT
2051 GAATAAGGGA TGTAACTACA AAAGCCCAT AGGGGGCAGT GTTTCGCCCG
2101 TGTTGTAGAA CTGGGTACAG AAAGGATCCT ATATGAAGTT CCTGAAACTG
2151 ACCTTTGTCT ATTATTACCT TCTCTGAAA GTGCCAGTCC ATGTATTTTT
2201 TATTATTTT AAGTTGTAAA TTTAATTTT AATTATTGTT ATGTGTTTGC
2251 ATTTAATTTT ATTTAATCAC CACATTTAGA AAATAATAAG AGCAAGTTTC
2301 TAAATGGGAG ACTGCTGAGG CTCTTTGCAA GAGATGAGAT TAAGTTTGAG
2351 TTTCTAAGGC AGGGCATGAG CTGGAATAG CATTGCTTTC CTTGATTGTC
2401 TCTCTCCTTC AGGGAGATTC TTTTCTCTA GTGTTTTAAG TGATCCTTTG
2451 AAGTAAGTGT GGAGAGTCTT GAATGGCAAG ACCAGGAGCT GAGTTTAAGC
2501 TTGTAATGGA AGCTTGCAAT GTGGGATATA TAACTGAGGA AGCATATTTA
2551 TCCTGAAGGT ATTTTGCCAG AAGGTATCAC TTGACCTGGA AAAGGAATCT
2601 ATTTAGTTCA GGAAGATATA AAGTTTAGA GGTATGTGAA GGAAGCACTT
2651 AGAACTTGCA AGCCTGATGT CCTATCAAGT TATGCTTCT GGGTGACAGA
2701 CAAAATAGCT TGTCTTATGG TGGTATGTTG TTGCATTTTC ACTTTGGGGT
```



```

2751 CTCTAAGAAA CTGTCAGTGA AAATATGTAC AATTCCTTCA ATTTCCATTG
2801 TTAACAACCTG TAATGTTGAA AAATAAGTTG AAAAGTCTTT GGGACCATAC
2851 ATGCAAAAAC GGTGCTCTG TTAATTAATT ATTTAATATT CTATAAATGT
2901 ACCCAATCTG TCCGCACCTT TCCAGTGAT GGGGAGATG GTCTGAGGAA
2951 GTATAATTTC AGTACTGGGG TCGGGAGAG GAGGTGATGT TTCTACATTT
3001 TTATTTTTTC TATAAATTGC AATTGGTCTG TATGCTGGTT TATTTTGAAA
3051 TTTATATTGG TTTCTTTTCA AGCTGGTGTG ATCTCCTAGA CTGTTTCACC
3101 CAGATGCTAG CATTTTTTTT TTTTGTAGA CAGAGTCTCA CTCTGTCACC
3151 TAGGCTGGAG TTGCAGTGGT TTGATCTCGG CTCACTGCAA CCTCCGACTC
3201 CTGGGTTCAG GCAATTCTTC TGCTCAGCC TCCTGAGTAG CTGGGATTAC
3251 AGATGTGCAC CAGCACACCC GGCTAATTTT TTGTATTTT AGTAGAGACA
3301 GGGTTTCGCC ATGTTGGCCA GGCTGGTCTT GAATCCTGG CCTTATGTGA
3351 TCCGCCACCC TTGGCTTCCC AAAGTGTGG GATTACAGGC ATGAGCCACC
3401 TCGCTGGGCC AGATGCTAGC ATTTTAGATC AAACAATTCA TTTTAGATGA
3451 ATTGTTTTGT TTCACAATCA TTTTAAATCA TTTTAGAATG TACTTCACAT
3501 TATTAGTTGT GTTATGGCAT AAAGGTACAA CCATTCCCTA ACTCCATCTT
3551 TTATTAATGC TTAAGTTTAA ATTATATTCT TCCATGCTT AAGCTATTCC
3601 CTAGAATTAA ACTGGGCACT TTGGAAGCA GCAACAGTAA CAGCAGCAGC
3651 AAACTTTTC TCTCATATT TGGGTGTATC AAAAGTTCTA GACTTTTGAA
3701 GTTATGATTT CAGTGGCCCA CTTTATTTCT AAGGAAGAGT GTCTACTTTG
3751 GAACGATACT TTGCACATAG TAGGAATCTA AGAATACAT TTGAATAATT
3801 ATAATTAAT GTTTAGCTAT CTTAATGAGA ATTTGTGAC AACAAAGAT
3851 CATCCATCGC CTTATGTGTG AGTAAGATTG GAGCCTCTAT CAAGATTTAG
3901 TCAAGTTCAG TTGATTTGAT TCTAGAAACA AATATTATT TCTTTCTTTT
3951 ACGGGGATGT GAATAAGGCT TTTCTTAAG GCCTTCATTC TTTAAACAAA
4001 CAGGTGAAA TGGTATGTTG TAAAAGAGAA GACGGGAGAG AGGTATTTAG
4051 ATGATAAGTG TACTTCACAA AAATGCCAAA GTTTGAAAAA TAGGTATGTT
4101 TGTCTAAAT GTTTAAGTGC TTCTCTGTTA GGTCTGGGG CTTCGAATCA
4151 TTTGAATTGT TCTGTTTTCAC AATAAAGGAG ATTCAGTGGG TTCTGCATTT
4201 TCAGGATTC AATAGAACTG TCCATTAATA AATAATCTCT TAGCAAGCAT
4251 TCGAATCCTA ACTGCTTTGA TGCATTGCC CTCGGGCACC TGTCATTTC
4301 AATATGGTAG GTGTCAAGT CAAAGTATT TACTGGGAGA AAAAGAGAG
4351 GAGTGGTTGT AGAAGTCTCC CTAATCAGA CATGTCAAGC AATCAGCCAA
4401 CGTGGTGTAT TTCTCATTTA ATATTTTAGT GTGAATTGAG ACACGTGAGT
4451 AAAGACATCG TGCAGAGATA AATGGGATA CAGTTAAATG TAGCAACTCT
4501 TGAGTTCAAT TTTTCCCACT GTAGCAAAAT TAATGCTTTC TCTTTATTGA
4551 AATAAATTGC TCATTCTTCC AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA
4601 AAAAAAGG

```

#### BLAST Results

Entry HSG27587 from database EMBL:  
human STS SHGC-32548.  
Score = 1951, P = 9.0e-101, identities = 411/425

Entry HS073350 from database EMBL:  
human STS EST303564.  
Score = 1417, P = 8.7e-58, identities = 285/287

#### Medline entries

No Medline entry

#### Peptide information for frame 2

ORF from the beginning to 580 bp; peptide length: 194  
Category: questionable ORF  
Classification: no clue

#### BLASTP hits

No BLASTP hits available

#### Alert BLASTP hits for DKF2phtes3\_18f3, frame 2

PIR:CGBOIS collagen alpha 1(I) chain - bovine (fragments), N = 1, Score = 155, P = 4.5e-10

TREMBL:HSCG1PAL1 gene: "COL1A1": Human proalpha 1 (I) chain of type I procollagen mRNA (partial)., N = 1, Score = 155, P = 6.5e-10

>PIR:CGBO1S collagen alpha 1(I) chain - bovine (fragments)  
Length = 779

## HSPs:

Score = 155 (23.3 bits), Expect = 4.5e-10, P = 4.5e-10  
Identities = 60/152 (39%), Positives = 67/152 (44%)

Query: 7 GEAGGPGAAWARRAAALPGTAA--GPPRPAAPPGA--APARGGPAPGAPQAALPRSQRGR 62  
G+ G PG + AR PG GPP PA P GA AP G A A P SQ  
Sbjct: 230 GDLGAPGPSGARGERGFPGERGVEGPPGAGPRGANGAPGNDGAKGDAGAPGAPGSQGAP 289

Query: 63 QLAERNRGRPRRRHGALAQPQGHGDLAAGVGRGAGGGHSRRGRHHVRSADLLQLPGAEE 122  
L G P RGA PG GD +GA G + G VR L + PG A  
Sbjct: 290 GL---QGMPGE-RGAAGLPGPKGDRGDAGPKGADGAPGKD---VRGLTGPIGPPGPAG 341

Query: 123 GAGDRGHL-P-GP-----DARDPELPRVFLPLAGLRGPPAA 156  
GD+G P GP D +P P P AG GPP A  
Sbjct: 342 APGDKGEAGPSGPAGTRGAPGDRGEPPG---P-AGFAGPPGA 381

Score = 121 (18.2 bits), Expect = 5.4e-05, P = 5.4e-05  
Identities = 52/154 (33%), Positives = 60/154 (38%)

Query: 7 GEAGGPGAAWARRAAALPGTAAGPPRPAAPPGAAPARG----GPAPGAPQAALPRSQRG 61  
G G PGAA R P AGPP P P G ++G GPA G P + P G  
Sbjct: 434 GATGFPGAA-GRVGGPPGSGNAGPPGPPGAGKEGSKGRGTGPA-GRPGEVGGPPGPPG 491

Query: 62 QLAERNRGRPRRRHGALAQPQGHGDLAAGVGRGAGGGHSRRGRHHVRSADLLQLPGA 121  
A G P G PG PG RG G +RG R L PG +  
Sbjct: 492 P--AGEKAGAGAD-GPAGAPGTGPGQIAGQGVVGLPGQRGE---RGFPGL---PGPS 541

Query: 122 EGAGDRGHLPGPDARDPELPRVFLPLAGLRGPPAAVRE 160  
G+G R P P + GL GPP + RE  
Sbjct: 542 GEPGKQGPSGASGERGPPGP---MGPPGLAGPPGESGRE 577

Score = 117 (17.6 bits), Expect = 1.8e-04, P = 1.8e-04  
Identities = 52/148 (35%), Positives = 62/148 (41%)

Query: 7 GEAGGPGAAWARRAAALPGTAAGPPRPA---PPGAAPARGGPAPGAPQAALPRSQRG-R 62  
G G PG AR +A PG A G P A PPG + GP PG P A +G R  
Sbjct: 416 GNVGAPGPKGARGSAGPPG-ATGPPGAAGRVGPPGPS-GNAGP-PPGPPGAGKEGSKGR 472

Query: 63 QLAERNRGRPRRRHGALAQPQGHGDLAAGVGRGAGGGHSRRGRH--HHVRSADLLQLPGA 120  
GRP G + PG PG GA G G + ++ LPG  
Sbjct: 473 GETGPAGRP---GEVGGPPGPAGEKAGAGADGPAGAPGTGPGQIAGQGVVGLPGQ 528

Query: 121 AEGAGDRGH--LPGPDARDPEL-PRVFLPLAGLRGPP 154  
G+RG LPGP + P +G RGPP  
Sbjct: 529 R---GERGFPLPGPSGEPGKQGPS---GASGERGPP 559

Score = 117 (17.6 bits), Expect = 1.8e-04, P = 1.8e-04  
Identities = 54/162 (33%), Positives = 64/162 (39%)

Query: 7 GEAGGPGAAWARRAAALPCT--AAGPPRPAAPPGAAPARG--GPA--PGAPQAALPRSQR 60  
G G PG + PG A+GP P PPG G G A PG P + P +  
Sbjct: 29 GPPGAPGPGQGPQGPGEPEGPGASGPMGPRGPPGPKNGDDGAGKPRGERGPPGPQ 88

Query: 61 G-QLAERNCRP--RRHGALAQPQGHGDLAAGVGRGAGGGHSRRGRHHV--RSLADLL 115  
G R L G P + HRG G GD +G G G + R L  
Sbjct: 89 GARGLPGTAGLPGMKHGRGFSGLDAGKADGAPGPKGEPGSPGENGAPGQMGPRGLPGFP 148

Query: 116 QLPGAA--EG-AGDRGHLPGPDARDPELPRVFLPLAGLRGPPAAA 157  
GAA G AG+RG +PGP P AG +GPP A  
Sbjct: 149 GPKGAAGEPKKAGERG-VGPPGAVG--PAGKDGEAGAGPPGPA 190

Score = 113 (17.0 bits), Expect = 5.4e-04, P = 5.4e-04  
Identities = 54/148 (36%), Positives = 58/148 (39%)

Query: 7 GEAGGPGAAWARRAAALPGTA-----AGPPRPAAP--PGAAPARGGPAP-GAPQAALPR 57  
G AG PGA A PG A AGPP PA P PG G P P GA A P  
Sbjct: 374 GFAGPPGADQPCAKGEPGDAGAKDAGPPGAPGAPGPIGNVAGPCKGARGSGAPP 433

Query: 58 SQRGRQLAERNRGRPRRRHGALAQPQGHGDLAAGVGRGAGGGHSRRGRHHVRSADLLQL 117  
G A P G PG PG +G G GR V  
Sbjct: 434 GATGFPGAAGRVGPPGSGNAGPPGPPGAGKEGSKGRGTGPAAGPGEVGP----- 486

Query: 118 PGAAGAGDRGHLPGPD--ARDPELPRVFLPLAGLRG 152  
PG AG++G PG D A P P +AG RG  
Sbjct: 487 PGPPGPAGEKG-APGADGPAGAPGTGPF-QGIAGQRG 521

Score = 110 (16.5 bits), Expect = 1.3e-03, P = 1.2e-03

Identities = 54/151 (35%), Positives = 60/151 (39%)

Query: 7 GEAGGPGAARAAALPGTAAGPPRPAAPPG--AAPAR-GGPAP-GAPALPRSRGR 62  
GE G A + L P G A G P A P G P G P P G A + +RG  
Sbjct: 194 GERGEQGPAGSPGFQGLPGA-GPPGEAGKPGEGVPGDLGAPGPSGARGERGFPGERGV 252

Query: 63 QLAERNRPRRRHRLAQPGHPGDLAAGVGRGAGGGHRRRHHHVRSLADLLQLPGAEE 122  
+ PR GA G GD A G+G +G R A L PG  
Sbjct: 253 EGPPGPAGPRGANGAPGNDGAKGDAGAPAGSGQAGPLQGMPE-RGAAGL---PGPK- 307

Query: 123 GAGDRGHLPGPDARD--PELPRVFLPLAGLRGPPAAA 157  
GDRG GP D P V L G GPP A  
Sbjct: 308 --GDRGDA-GPKGADGAPKGDV-RGLTGPIGPPGPA 340

Score = 109 (16.4 bits), Expect = 1.7e-03, P = 1.7e-03  
Identities = 55/154 (35%), Positives = 60/154 (38%)

Query: 4 NGN-GEAGGPGAARAAALPGTAAGPPRPAAPPGAAPARG-GPAPGAPALPRSRGR 61  
NG+ GEAG PG R P A G P A PG RG GA A P +G  
Sbjct: 67 NGDDGEAGKPRRP-GERGPPPGQARGLPAGTGLPMKGRGFSGLDGAAGDAGPAGPKG 125

Query: 62 RLAE-RNGRPRRRHRLAQPGHPGDLAAGVGRGAGGGHRRRHHHVRSL----ADLL 115  
+ NG P + G PG PG A G G V A  
Sbjct: 126 EPGSPGENGAPGQ-MGPRGLPGFPKGAAGEPGKAGERGVPGPPGAVGPAKDGEGAQ 184

Query: 116 QLPGAAGAGDRGHLPGPDARDPELPRVFLPLAGLRGPPAAA 157  
PG A AG+RG GP A P F L G GPP A  
Sbjct: 185 GPPGPAGPAGERGE-QGP-AGSPG---FQGLPGFAGPPGEA 220

Score = 104 (15.6 bits), Expect = 6.6e-03, P = 6.6e-03  
Identities = 44/131 (33%), Positives = 49/131 (37%)

Query: 2 EVNGNGEAGGPGAARAAALPGTAAGPPRPAAPPGAAPARGGPAP-GAPALPRSRGR 60  
E GE G PG R L P G GP A PG A RG P P G A A +  
Sbjct: 126 EPGSPGENGAPGQMGPR---GLPGFP-GPKGAAGEPGKAGERGVPGPPGAVGPAKDGEGA 181

Query: 61 GRQLAERNRPRRRHRLAQPGHPGDLAAGVGRGAGGGHRRRHHHVRSLADLLQLPGA 120  
G Q P RG G PG G+ G G G+ DL PG  
Sbjct: 182 GAQGPFPAGPAGERGEQGPAGSPG--FQGLP-GPAGPGEAGKPGEGVPGDL-GAPGP 237

Query: 121 AEGAGDRGHLPG 132  
+ G+RG PG  
Sbjct: 238 SGARGERG-FPG 248

Score = 104 (15.6 bits), Expect = 6.6e-03, P = 6.6e-03  
Identities = 43/131 (32%), Positives = 55/131 (41%)

Query: 7 GEAGGPGAARAAALPGTAAGPPRPAAPPGAAPARGGPAPGAPALPRSRGRQLAE 66  
GEAG G A R A PG G P P P G A ARG GP PGA Q + + G A+  
Sbjct: 347 GEAGPSGPAGTRCA---PGDR-GEPPPGPAGFA---GP-PGADGQPGKAGEPGDAGAK 397

Query: 67 RNRGRPRRRHRLAQPGHPGDLAAGVGRGAGGGHRRRHHHVRSLADLLQLPGAEEGAGD 126  
+ P G PG G+ A +GA G G + A + PG + AG  
Sbjct: 398 GDAGPPGPAGPAGPPGPIGNVGAAPGPKGARGSGATGFPGA-AGRVGPPGPSGNAGP 456

Query: 127 RGLPGPDARD 137  
G PGP ++  
Sbjct: 457 PGP-PGPAGKE 466

Score = 104 (15.6 bits), Expect = 6.6e-03, P = 6.6e-03  
Identities = 56/162 (34%), Positives = 62/162 (38%)

Query: 7 GEAGGPGAARAAALPGTAA--GPPRPAAPPGAAPARGGPAPGAPALPRSRGRQL 64  
G G PGA A G GP P P G A ARG P P Q PR +G  
Sbjct: 608 GPPGAPGAPGPVGPAGKSGDRGETGPAGPIGPVGPAGARG---PAGP-QG-PRBKGZTG 662

Query: 65 AERNRPRRRHRG---ALAQPGHPGDLAAGVGRGAGGGHRRRHHHVRSLA-DLLQ-LPG 119  
+ + + HRG PG PG GA G RG S D L LPG  
Sbjct: 663 ZGBRGIKGHRGFSGLQPPGPPGPSGEQGPSGASGPAGPRGPPGSAGSPGKDLNGLPG 722

Query: 120 AAGAGDRGHL--PGPDARDPELPRVFLPLAGLRGPPAAVREERLHRPVQ 168  
G RG GP A P P P G GPP+ L +P Q  
Sbjct: 723 FIGPPGPRGRTGDAGP-AGPPGPPG---P-PGPPGPPSGGYDLSFLPQPPQ 768

Score = 101 (15.2 bits), Expect = 1.5e-02, P = 1.5e-02  
Identities = 49/148 (33%), Positives = 55/148 (37%)

Query: 7 GEAGGPGAARAAALPGTAAGPPRPAAPPGAAPARGGPAPGAPAAA--QALPRSRGR 62  
G AG PG A R PG A GP A G A+G P P PA + P G  
Sbjct: 152 GAAGEPGKAGERGVPGPPG-AVGP---AGKDGEAGAQPPGPAGPAGERGEQGPAGSPGF 207

Query: 63 QLAERNRPRRRHRLAQPCHPGDLAAGVGRGAGGGHSRRGRHHVRSADLLQLPGAEE 122  
Q P G + G PGDL A G G RG R + PG A  
Sbjct: 208 QGLPGPAGPPGEAGKPGEQGVPGDLGAP---GPSGARGERGFPGE-RGVEGP---PGPAG 260

Query: 123 GAGDRGHLPGPDARDPELPRVFLPLAGLRGPP 154  
G G PG D + P G +G P  
Sbjct: 261 PRGANG-APGNDGAKGDAGAPGAP--GSQGAP 289

Score = 100 (15.0 bits), Expect = 1.9e-02, P = 1.9e-02  
Identities = 40/130 (30%), Positives = 48/130 (36%)

Query: 7 GEAGGPGAARAAALPGT---AAGPPRPAAPPGAAPARG--GPA--PGAPAQALPRSQR 60  
G G PG + PG A+GP P PPG G G A PG P + P +  
Sbjct: 29 GPPGAPGPGFQGPGEPEPGASGPMGPRGPPGPGKNGDDGEAGKPRPGERGPPGPQ 88

Query: 61 G-RQLAERNRPR--RRHRGALAQPCHPGDLAAGVGRGAGGGHSRRGRHHVRSADLLQL 117  
G R L G P + HRG G GD +G G G + L  
Sbjct: 89 GARGLPGTAGLPGMKHGRGFSGLDGAAGDAGPAGPKGEPGSPGENGAPQMGPRG-LPGF 147

Query: 118 PGAAEGAGDRG 128  
PG AG+ G  
Sbjct: 148 PGPKGAAGEPG 158

Score = 99 (14.9 bits), Expect = 2.5e-02, P = 2.5e-02  
Identities = 53/156 (33%), Positives = 61/156 (39%)

Query: 7 GEAGGPGAARAAALPGT---AAGPPRPAAPPGAAPARG--GPA---PGAPAQAL 55  
G G PGA R A PG A G P P P G + RG GPA P PA A  
Sbjct: 587 GRGSPGAKGDRGETGPAGAPGPPGAPGAPGPGVPGAGKSGDRGETGPAGPIGPVGPAGAR 646

Query: 56 -----PRSQGRQLAERNRPRRRHRLAQPCHPGDLAAGVGRGAGGGHSRRGRHHV 108  
PR +G + + + HRG G PG + +G G G  
Sbjct: 647 GPAGPQGPGBKGZTGZGZGBRGIKHGRGFSGLQGPFGPGSPGEQGPSGASGPAGPRGP- 705

Query: 109 RSLADLLQLPGAAGAGDRG--HLPGPDAARDPELPRVFLPLAGLRGPP 154  
PG+A G G LPGP P PR AG GPP  
Sbjct: 706 -----PGSAGSPGKDLNGLPGPIG--PPGPRGRTGDAGPAGPP 742

Score = 98 (14.7 bits), Expect = 3.3e-02, P = 3.3e-02  
Identities = 51/158 (32%), Positives = 58/158 (36%)

Query: 7 GEAGGPGAARAAALPGTA-----AGPPRPAAPPGAAPARGGPAP-GAPAQALPRSQR 60  
G G G R AA LPG AGP PG RG P G P A +  
Sbjct: 287 GAPGLQGMGERGAAGLPGPKGDRGDAGPKGADGAPGKDGVRGLTGPIGPFGPAGAPGDK 346

Query: 61 GRQLAERNRPRRRHRLAQPCHPGDLAAGVGRGAGGGHSRRGRHHVRSADLLQL 117  
G A +G P RGA +PG PG + D GA G +G + D  
Sbjct: 347 GE--AGPSG-PAGTRGAPGDRGEPPGPPGAGFAGPPGADGQPAKGEPEGDAGAKGDAGP- 402

Query: 118 PGAAEGAGDRGHLPGPDARDPELPRVFLPLAGLRGPPAAVR 159  
PG A AG G + A P+ R G G P AA R  
Sbjct: 403 GPPAGPAGPPGPIGNVGAAGPKGARGSGPPGATGPPGAAGR 444

Score = 96 (14.4 bits), Expect = 5.7e-02, P = 5.5e-02  
Identities = 46/152 (30%), Positives = 57/152 (37%)

Query: 6 NGEAGGPGAARAAALPGTAA---GPPRPAAPPGAAPARGGPAPGAPPA-QALPRSQRGR 62  
+G G PGA + PG G PA PG A G P P PA ++ R + G  
Sbjct: 574 SGREGAPGAEGSPGRDGSAGKGDGETGPAGAPGPPGAPGAPGVPAGKSGDRGETGP 633

Query: 63 QLAERNRPRRRHRLAQPCHPGDLAAGVGRGAGGGHSRRGRHHVRSADLLQLPGAEE 122  
P RG G G+ +G G RG HR + L PG  
Sbjct: 634 AGPIGPVGPAGARGPAGPQGPGB-----KGZTGZGZGBRGIKH-RGFSGLQGPFGPPG 686

Query: 123 GAGDRGHLPGPDARDPELPRVFLPLAGLRGPPAAA 157  
G++G P A P AG RGPP +A  
Sbjct: 687 SPGEQG--PS-GASGP-----AGPRGPPGSA 709

Score = 94 (14.1 bits), Expect = 9.7e-02, P = 9.2e-02  
Identities = 45/134 (33%), Positives = 56/134 (41%)

Query: 24 PGTAAGPPRPAAPPGAAPARGGPA-PGAPAQALPRSQRQLAERNRPRRRH--GALAQ 80  
P G P P PG +G P PG P + P RG G P ++ G +  
Sbjct: 21 PSGRPLGPPGAPGPGQGPGEPEPGASGPMGPRGPP-----GPPGKNGDDGEAGK 75

Query: 81 PGHPGDLAA-GV--GRGAGGGHSRRGRHHVRSADLLQLPGAAGAGDRGH--LPGPDA 135  
PG PG+ G RG G G HR + L GA AG +G PG +  
Sbjct: 76 GPRPGERGPPGPGARGLPGTAGLPGMKGH-RGFSGLDGAAGDAGPAGPKGEPGSPGENG 134

Query: 136 RDEL-PRVFLPLAGLRGPPAAA 157  
++ PR LP G GP AA

Sbjct: 135 APQMGPGRG-LP--GFPGKGAA 154

Score = 92 (13.8 bits), Expect = 1.7e-01, P = 1.5e-01  
Identities = 52/155 (33%), Positives = 58/155 (37%)

Query: 7 GEAGGPGAAWARRAAALPGTAAGPPRPAAPPGAAPARGGP-APGAPALPRSQGRQLA 65  
GEAG G A R A G GPP PA G A G P A G P A + G

Sbjct: 347 GEAGPSGPAGTRGAPGDRGEP-GPPGPAGFAGPPGADGQPGAKGEPGDAGAKGAGPPGP 405

Query: 66 ERNGRPRRRHAGALAQPGHPGDLAAGVGRGAGGGHSRRGR--HHVRSIADLLQLPGA-- 121  
P G + PG G + GA G GR A PG A

Sbjct: 406 AGPAGPPGPIGNVGAAPGPKGARGSAGPPGATGFPGAAGRVGPPGPSNAGPPGPPGPAGK 465

Query: 122 EGA-GDRGHLPGPDARDPELPRVFLP-LAGLRGPPAA 156  
EG+ G RG GP R E+ P AG +G P A

Sbjct: 466 EGSKGPRGET-GPAGRPGEVGPFGPPGAGEKAGPA 501

Score = 92 (13.8 bits), Expect = 1.7e-01, P = 1.5e-01  
Identities = 51/156 (32%), Positives = 57/156 (36%)

Query: 7 GEAGGPGAAWARRA---AALPCT--AAGPPRPAAPPGAAPARGGPAPGAPAL-PRSQ 60  
G G PGA R A PG A G P P P G + RG P P + P R

Sbjct: 587 GRDGSFGAKGDRGETGPAGAPGPPGAPGAPGVGPKSGDRGETGPAGPIGPVGPAGAR 646

Query: 61 GRQLAERNRPRRRHAGALAQPGHPGDLA-AGVG--RGAGGHSRRGRH--HHVRSIADLL 115  
G A G PR +G + G G G +G G G A

Sbjct: 647 GP--AGPQG-PRBGKZTGZGZGBRGKIGHRGFSGLQGGPPGPSGPGSGASGPAGPR 703

Query: 116 QLPGAAGAGDRG--HLPGPDARDPELPRVFLPLAGLRGPP 154  
PG+A G G LPPG P PR AG GPP

Sbjct: 704 GPPGSAGSPGKDLNGLPGPTG--PPGPRGRTGDAGAPGPP 742

Score = 90 (13.5 bits), Expect = 2.8e-01, P = 2.5e-01  
Identities = 45/134 (33%), Positives = 53/134 (39%)

Query: 7 GEAGGPGAAWARRAAALPGTAAGPPRPAAPPGAAPARGGPAPGAPALPRSQGRQ-LA 65  
G G PG A + A G A P P P G A RG G P Q R +RG L

Sbjct: 485 GPPGPPGPAGEKAGAPGADGPAGAPGTPG-PQGIAGQRG--VVGLPGQ--RGERGFPL 538

Query: 66 ERNGRPRRH--RGALAQPGHPGDLA---AGV---GR-GAGGHSRRGRHHVRSIADL 114  
+G P + GA + G PG + AG + D GR GA G GR +

Sbjct: 539 GPSGEPGKQGPSGASGERGPPGMPGPPGLAGPPGESGREGAPGAEGSPGRDGSFGAKGDR 598

Query: 115 LQL-PGAAEGAGDRGHLPGP 133  
+ P A G PGP

Sbjct: 599 GETGPAGAPGPPGAPGAPG 618

Score = 83 (12.5 bits), Expect = 1.8e+00, P = 8.3e-01  
Identities = 49/156 (31%), Positives = 56/156 (35%)

Query: 7 GEAGGPGAAWARRAAALPGTAAGPPRPAAPPGAAPARGGPAPGAPALPRSQGRQ-LA 65  
C+AG GA A + G GPP PA PG G GFA GAP R +

Sbjct: 311 GDAGPKGADGAPGKDGVRGLTGPIGPPGAPGDKGACGSPGAGTRGAPGD---RCEP 367

Query: 61 GRQLAERNRPRRRHAGALAQPGHPGDLAAGVGRGAGGGHSRRGRHHVRSIADLLQLPGA 120  
P G G PGD A G G G + ++ PG

Sbjct: 368 GPPGPAGFAGPPGADGQPGAKGEPDGAGKDGAGPPGAPGAPGPPGPIGNV----APGP 423

Query: 121 AEGAGDRGHLPGPDARDPELPRVFLP----LAGLRGPPAAVRE 160  
G G PG RV P AG GPP A +E

Sbjct: 424 KGARGSAGP-PGATGFPGAAGRVGPPGPSNAGPPGPPGPAGKE 466

Score = 82 (12.3 bits), Expect = 2.3e+00, P = 9.0e-01  
Identities = 46/148 (31%), Positives = 52/148 (35%)

Query: 7 GEAGGPGAAWARRAAALPGTAAGPPRPAAPPGAAPARGGPAPGAPALPRSQGRQLAE 66  
G+AG PGA ++ A LG G A PG RG P A P RL

Sbjct: 275 GDAGAPGAPGSQAPGLQGMPP-GERGAAGLPKPGKDRGDAGPKG-ADGAPKDGVRGLTG 332

Query: 67 RNRGRPRRRHAGALAQPGHPGDLAAGVGRGAGGGHSRRGRHHVRSIADLLQLPGAAGAGD 126  
G P G PG G+ G G RG A PGA G

Sbjct: 333 PIGPP---GPAGAPGDKGAGPSGPAGTRGAPGDRGEPGPPGP-AGFAGPPGADGQPGA 387

Query: 127 RGHLPGP-DARDPELPRVFLPLAGLRGPP 154  
+G PG A+ P P AG GPP

Sbjct: 388 KGE-PGDAGAKGDAGPPG--P-AGPAGPP 412

Peptide information for frame 3

-----

ORF from 12 bp to 755 bp; peptide length: 248  
 Category: similarity to known protein  
 Classification: unset  
 Prosite motifs: LEUCINE\_ZIPPER (17-39)  
 LEUCINE\_ZIPPER (24-46)

## BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphtes3\_18f3, frame 3

TREMBL:AF070675.1 product: "TNF-inducible protein CG12-1"; Homo sapiens TNF-inducible protein CG12-1 mRNA, complete cds., N = 1, Score = 135, P = 1e-06

TREMBL:HS6802.1 gene: "dJ6802.1"; product: "dJ6802.1"; Homo sapiens DNA sequence from PAC 6802 on chromosome 22. Contains apolipoprotein L, myosin heavy chain, ESTs, CA repeat, STS and GSS., N = 1, Score = 107, P = 0.0023

>TREMBL:AF070675.1 product: "TNF-inducible protein CG12-1"; Homo sapiens TNF-inducible protein CG12-1 mRNA, complete cds.  
 Length = 331

## HSPs:

Score = 135 (20.3 bits), Expect = 1.0e-06, P = 1.0e-06  
 Identities = 30/103 (29%), Positives = 55/103 (53%)

Query: 30 RLHRQVLRRLREVARRLERLRRRLVANVAGSSLSATGALAAIVGLSLSPVTLGTSLVSA 89  
 ++ + +LR +A +E + R ++NV SS A + ++ GL L+P T GTSL ++A  
 Sbjct: 91 KIQESIEKLRALANGIEEVHRCCTISNVVSSSTGAASGIMSLAGLVLPFTAGTSLALTA 150

Query: 90 VGLGVATAGGAVTITSDL-SLIFCNSRELRRVQEIAATCQDQMR 132  
 G+G+ A IT+ + + +S E + AT D+++  
 Sbjct: 151 AGVGLGAASAVTGITTSIVEHSYSSAEAE-ASRLTATSIDRLK 193

Pedant information for DKFZphtes3\_18f3, frame 2

## Report for DKFZphtes3\_18f3.2

{LENGTH}	193
[MW]	19708.24
[pI]	11.90
[KW]	All_Alpha
[KW]	LOW_COMPLEXITY 55.44 %

SEQ TEVNGNGEAGGPGAAMARRAAALPGTAAGPPRPAAPPGAAPARGGPAPGAPQAQALPRSOR  
 SEG .....XXX...  
 PRD cccccccccccccchhhhhhhhhccccccccccccccccccccccccccccccccchhhhhh

SEQ GRQLAERNRPRRRHRCALAQPGHFGDLAAGVGRGAGGGHSRRGRHHVRSADLLQLPGA  
 SEG .....XXXXXXXXXXXXXXXXXXXXXXXXXXXXX...  
 PRD hhhhhhccchhhhhhhhhcccccc

SEQ AEGAGDRGHLPGPDARDPELPRVFLPLAGLRGPPAAVREERLHRPVQFCLLHRLWLWTW  
 SEG .....XXXXXXXXXXXXX...  
 PRD cchhhhhhhccccchhhhhhhhhhhc

SEQ LPHPGAGGGGHQ  
 SEG XXXXXXXXXXXXX  
 PRD ccccccccccccc

(No Prosite data available for DKFZphtes3\_18f3.2)

(No Pfam data available for DKFZphtes3\_18f3.2)

Pedant information for DKFZphtes3\_18f3, frame 3

## Report for DKFZphtes3\_18f3.3

{LENGTH} 248  
 {MW} 27162.56  
 {pI} 9.92  
 {PROSITE} LEUCINE\_ZIPPER 2  
 {KW} TRANSMEMBRANE 1  
 {KW} LOW\_COMPLEXITY 30.65 %  
 {KW} COILED\_COIL 12.10 %

SEQ MGERPAAREPHGPDALRRFQGLLDRRGRLHRQVLRLEVARRLRLRRSLVANVAGS  
 SEG .....XXX  
 PRD cccccccccccccchhhcc  
 COILS .....  
 MEM .....

SEQ SLSATGALAAIVGLSLSPVTLGTSLLVSAVGLGVATAGGAVTITSLSLIFCNSRELRRV  
 SEG xxxxxxxxx.....XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX  
 PRD cchhhhhhhhhhhhhccccccccccccccccccccccccccccccccccccccchhhhhhh  
 COILS .....  
 MEM .....MMMMMMMMMMMMMM.....

SEQ QEIAATCQDQMRILSCLEFFCRWQCGDRQLQCGRNASTALYNSVYFIVFSGSRGFLI  
 SEG .....  
 PRD hhhcccccccccccccccc  
 COILS .....  
 MEM .....

SEQ PRRAEGDTKVSQAVLKAKIQKLAESLESCTGALDELSEQLSRVQLCTKSSRGHDLKISA  
 SEG .....  
 PRD cccccccccchhhcccccccccccc  
 COILS .....CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC  
 MEM .....

SEQ DQRAGLFF  
 SEG .....  
 PRD hhhhhccc  
 COILS .....  
 MEM .....

## Prosite for DKFZphtes3\_18f3.3

PS00029	17->39	LEUCINE_ZIPPER	PDOC00029
PS00029	24->46	LEUCINE_ZIPPER	PDOC00029

(No Pfam data available for DKFZphtes3\_18f3.3)

DKFZphtes3\_1817  
-----

group: cell structure and motility

DKFZphtes3\_1817 encodes a novel 1050 amino acid protein with weak partial similarity to ankyrins.

The novel protein contains an ATP/GTP-binding site motif A (P-loop) and an Ank repeat. Ankyrins are peripheral membrane proteins which interconnect integral proteins with the spectrin-based membrane skeleton. Thus the novel protein seems to be involved in coupling of cyto skeleton and cell membrane.

The new protein can find application in modulation of cyto skeleton-membrane interactions.

similarity to ankyrins

Sequenced by MediGenomix

Locus: unknown

Insert length: 4501 bp

Poly A stretch at pos. 4423, no polyadenylation signal found

```
1 GATCGCCGCG CGAGGGTGGT GGGCATCGAG GTCCCAGCAG CGGACGAGGG
51 AGGTGCCGCC GTCGCCCAGG ATGGGCTGGG AATGAAGCGA TGTAGCCTTT
101 TAAGAGATTG GCTCTGACCC ATCTGAAGTC CATATGGCTC TGTATGATGA
151 AGACCTCCTG AAAAAATCCTT TCTATCTGGC TCTGCAAAAG TGCCGCCCTG
201 ACTTGTGTCG CAAAGTGGCC CAAATCCATG GCATTGTCTT AGTACCCCTG
251 AAAGGAAGCC TGTGAGCAGC CATCCAGTCT ACTTGTCACT TTGAGTCCCTA
301 CATTTTGATA CCTGTGGAGG AGCATTTCAT GACCTTAAAT GGAAGAGATG
351 TCTTTATTCA AGGGAACAGG ATTAATTAG GAGCTGTTT TGCCCTGTCTT
401 CTCTCAGTGC CCATTCTCTT TGAAGAAACT TTCTACAATG AGAAGAGAGA
451 GAGTTTCAGC ATCCTGTGTA TAGCCCATCC TTGGAAGAG TTGAGAGAGT
501 CAGAAGAGCC TTGGGCACCC TCAGATCCCT TTCCCTGAA AACCAATGAA
551 GATGTGAGAG AGTTCTTGGG AAGACACTCC GAGGCAATTG ACAGGAACAT
601 CGCCTCTTTC CATCGACCAT CCTCTTACA CCAATCCCT CCAGCAGCTT
651 ACCACATAGA CTCACCTGAA AATCCTCCG AAGCAGAGG CCCAGATGAA
701 CTGAGGAGCT CTCAGGCAAT CCTCTTACA CCAATCCCT CCAGCAGCTT
751 CCTGATGAAG CAGGCAATG AGATATACCT CCATCATGAA ATTTACAAAC
801 TGATCTTTAA ATACGTGGG ACCATGGAG CAAGTGAGGA TCGGCGCTTT
851 AACAAATCA CAAGAAGCCT TCAAGATCTT CAGCAGAAAG ATATTGGTGT
901 GAAACCGGAG TTCAGCTTTA ACATACCTCG TGCCAAAGA GAGCTGGTCT
951 AGCTGAACAA ATGCACCTCC CCACAGCAGA AGCTTGTCTG CTTCGCAAAA
1001 GTGGTGCAGC TCATTACACA GTCTCAAGC CAGAGAGTGA ACCTGGAGAG
1051 CATGTGTGCT GATGATCTGC TATCAGTCTT GTTATACCTG CTGTGTAAGA
1101 CGGAGATCCC TAATTGGATG GCAAAATTGA GTTACATCAA AAACCTCAGG
1151 TTTAGCAGCT TGCCAAAGGA TGAAGTGGGA TACTGCTGTA CCTCATTCGA
1201 AGCTGCCATT GAATATATTC GGCAAGGAAG CCTCTCTGCT AAACCCCTG
1251 AGCTGAGGGG ATTTGGAGAG AGGCTGTCTT TTAAGCAGAG AATGAGCTTA
1301 CTCTCTCAGA TGACTTCGTC TCCACCCGAC TGCGTGTITA AGCACATTGC
1351 ATCAGGTAA CAGAAAGAAG TGGAGAGACT TCTGAGCCAA GAGGACCATG
1401 ATAAAGATAC CGTCCAAAAG ATGTGTACCC CTCTGTGCTT CTGCGATGAC
1451 TGTGAGAAAC TCGTCTCTGG GAGGTGTAAT GATCCCTCAG TTGTCACTCC
1501 ATTTCTCAGA GACGACAGGG GGCACACCCC TCTCCATGTG CGTGTGTCT
1551 GTGGGCGAGG ATCCCTCATC GACCTCCTGG TTCCAAAGGG CGCCATGGTA
1601 AATGCCACAG ACTACCATGG GGCCACTCCG CTCACCTGGG CCTGTGAGAA
1651 GGGCTACCAG AGCGTGACGC TGCTGTGCTT GCACCTAAGG GCCAGCGCGG
1701 AAGTGCAGGA CAACAATGGG AATACGCCAC TCCACTGGG CTGACCTAC
1751 GGCACGAGG ACTGTGTGAA GGCTCTGTTT TACTACGAGC TGGAGTCTGT
1801 CAGACTTGAC ATTGGCAATG AGAAAGGAGA CACCCTCTCA CACATTGCTG
1851 CCGCTGGGG CTACCAAGGC GTCATAGAGA CATTGTGCA GAACGGAGCG
1901 TCCACCGAGA TCCAGAACAG ACTGAAGGAG ACGCCCTCA AGTGTGCATT
1951 AAACCTCAAG ATTCTGTCTG TAATGGAAGC CTATCACCTG TCCTTCGAGA
2001 GGAGGCAGAA GTCGTCCGAG GCCCTGTGTC AGTCCCGGCA GCGCTCCGTG
2051 GACTCCATCA GCCAAGATGC CTCCACTTCC AGCTTCTCCT CCATGTGAGC
2101 CGGCTCAAAG CAGGAGGAGA CCAAGAAGGA CTACAGAGAG GTAGAAAAAC
2151 TTTTGAGAGC AGTTGTCTGAT GGAGATCTAG AAATGGTGCG TTACTTGTG
2201 GAATGGACAG AGGAGGACCT GGAGGATGCG GAGGACACTG TCAGTGCAGC
2251 AGACCCGAA TTCTGTACCC CGTTGTGCCA GTGCCCAGG TGTGCCCCAG
2301 CTCAGAAGAG GCTGGCGAAG GTTCTTGCCA GTGGGCTTGG TGTGAACGTG
2351 ACCAGCCAGG ACGGCTCTCT CCCGCTGCAT GTGCGGCCCC TGACGGCCCG
2401 GCGGACCTC ATCCGCTCTC TGCTGAAGCA GGGGGCCAAC GCAGGTGCCA
2451 GGAACGCAGA CCAAGCCGTC CCGCTCCACC TGGCCTGCCA GCAGGGCCAC
2501 TTTTCAAGTG TGAAGTGTCT GTTAGATTCT AATGCAAAAC CCAATAAGAA
2551 GGACCTCAGT GGAACACGCG CCCTCATTTA CGCTGCTCT GGTGGCCATC
2601 ACGAGCTTGT GGCACTGCTG CTACAGCAGC GGGCTCCAT TAACGCTTCT
2651 AACAAATAGG GCAACACAGC GCTGCACGAG GCTGTGATTG AAAAGCACGT
```



```

2701 CTTCTGGTGA GAGCTGCTTC TGCTCCACGG AGCGTCAGTT CAGGTGCTGA
2751 ACAAGCGGCA GCGCAGGGCT GTAGACTGTG CTGAACAGAA TTCAAAAAATA
2801 ATGGAATTGC TTCAGGTGGT ACCAAGCTGT GTTGCTTCAT TAGATGATGT
2851 GGCTGAAACT GACCGCAAGG AGTATGTCAC TGTTAAGATC AGGAAAAAAT
2901 GGAACCTCAA ACTGTATGAT CTACCAGATG AGCCTTTTAC AAGACAGTTT
2951 TACTTTGTCC ACTCAGCTGG TCAGTTTAAG GGAAAGACTT CAAGGGAGAT
3001 TATGGCAAGA GATAGAAGTG TCCCTAATTT AACCGAAGGT TCTTTGCATG
3051 AGCCGAGGAG GCAAAGTGTC ACACCTGAGC AGAATAACCT GCCAGCTCAG
3101 AGTGGATCTC ATGCTGCTGA GAAAGGCAAC AGCGACTGGC CAGAGAGGCC
3151 TGGACTGACA CAGACTGGCC CTGGACACAG ACAGGATGCT CGGAGACACA
3201 CGGTAGAGGA TGGCGTCGTG TCCGAGGCC CGGAGGCTGC TGGCCCCCTC
3251 TCCACTCCCC AAGAGGTTAG TGCTCCCGG TCCTAACAGG AATGAGGAGT
3301 TGTGTAACCC ACTGCTAGGA AGCAAGGATG CAACAAGATG ATGCTGAGCG
3351 TGAACACATC TGAGAACTAA ATGTGCTTCC ATGAGACTGG CTTGAGAAGT
3401 CTTCAGCACC AAGTTCCTGA AAGCTTTTCT GTGGCAGGAA AGAATGCAAC
3451 AAAAAAGTTA ACCACCACCA TCTCTCTCTT CTTCAAAGCT AATGAATACA
3501 ATTGAAACAG ACAAAAATTC CAGTAGCATC CAGATCCTTA AGCCAGAGGT
3551 GCATGCTTCT TTTTAAGTAT GAGGGTTTGT TGGTCACAGT GGGAGAGGTT
3601 TCACCACCGC ATTCTGACCT CCTCCTCCCA AAAGGTGCTA AACCTCTCTG
3651 ACCTGTGTAC ATTCACAAAC CACAGCTAGA ATTCTCCAC CTAGGATTAA
3701 GCTGGAGAGA AGTAAGTAAT TTAGGTTTCA TGGTACTGTA GAGGCCAGGC
3751 TGAATGTGTA TATCTGAAGG AAGAAAGCAG CAGCTGGACA ATGTTTCTTT
3801 GCAAAGCAAC ACTCGAACCA AAAGATGCCT CAATCCCAT TGTATATTCA
3851 TTTTAGTGAA AGGATGCATC AGACCTGTTC CACATCATGC ACATGGGAAA
3901 GGGTGGTTAT CATTTTCCTT CTAACAAGTA GGTACAGATA TTCGGTTACT
3951 ACACGTGCAC CTGTAGCAGT ATTTCTAGAA ACATCCCTTT TTGTTGAGAA
4001 CCTCCCTTGA ATGTCTGTCA CACTCACACC TGACGGGATG GTTACTGGAT
4051 TAGAGAGTAG ATTTGGCACA TCTTTTCTTA GTCTTTTAT TCAAAATCAA
4101 AACTTAACAG CACAAACCAG GTCAGAGTTA CTTTCGGTTA GAATTTATTG
4151 CCATTTATTC CTTTTTATAA ATTTCTATAG ATTATCTGT TATTTTATG
4201 TTATTGGCCT AGAGCTACAC GTATATGGGT TTGCTCTGAG TCCGTTTCA
4251 AATGACCTTG TGATAGGAAA ATGGTTTGT CCATGTTCTT GGAAATACTT
4301 GTGTATGTAC AGAAGGAAGG GAGGGATTAT TTTCTACAA AGTAATTTAT
4351 GATTTCATAA TTTCTAATGT GCCTTGGATA TGTGCCAAAT GATGGAAAAG
4401 AAACAGTAAA CTTTATGATT CTTAAAAAAA AAAAAAATAA AAAAAAATAA
4451 AAAAAAATAA AAAAAAATAA AAAAAAATAA AAAAAAATAA AAAAAAATAA
4501 G

```

## BLAST Results

No BLAST result

## Medline entries

No Medline entry

## Peptide information for frame 2

ORF from 134 bp to 3293 bp: peptide length: 1050  
 Category: similarity to known protein  
 Classification: Cell structure/motility  
 Prosite motifs: ATP\_GTP\_A (945-953)

```

1 MALYDEDLLK NPFYLALQKC RPDLSKVAQ IHGIVLVPCK GSLSSSIQST
51 QCFESYILIP VEEHFQTLNG KDVFIQGNRI KLGAGFACLL SVPILFEETF
101 YNEKEESFSI LCIAHPLEKR ESSEELAPS DPFLSKTIED VREFLGRHSE
151 RFDNRNIAFSH RTFRECEKRS LRHHIDSANA LYTKCLQOLL RDSHLKMLAK
201 QEAQNLMLKQ AVEIYVHHEI YNLIFKYVGT MEASEDAAFN KITRSLQDLQ
251 OKDIGVKPEF SFNIPRAKRE LAQLNKCTSP QOKLVCLRVK VOLITQSPSQ
301 RVNLETMCAD DLLSVLLYLL VKTEIPNWMN NLSYIKNFRF SSLAKDELGY
351 CLTSFEAAIE YIROGSLSAK PPESEGFQDR LFLKORMSLL SQMTSSPTDC
401 LFKHIASGNQ KEVERLLSQE DHDKDTVQKM CHPLCFDDC EKLVSGRIND
451 PSVVTFFSRD DRGHTPLHVA AVCGQASLID LLVSKGAMVN ATDYHGATPL
501 HLACQKGYOS VTLLLLHYKA SAEVODNNGN TPLHLACTYG HEDCVKALVY
551 YDVESCRLOI GNEKGDTPH IAAWGYQGV IETLLONGAS TEIQNRLKET
601 PLKCALNSKI LSVMEAYHLS FERRQKSSEA PVQSPQRSVD SISQESSTSS
651 FSSMSAGSRQ EETKDYREV EKLLRAVADG DLEMVRYLLE WTEEDLEDAE
701 DTVSAADPEF CHPLCQPKC APAQKRLAKV PASGLGVNVT SQDGSPLPHV
751 AALHGRADLI RLLKKGANA GARNADQAVP LHLACQGHF QVVKCLLDSN
801 AKPNKKDLGS NPLIYACSG GHHELVALLL QHGASINASN NKGNTALHEA
851 VIEKHVFVVE LLLHGHASVQ VLNKRQRTAV DCAEQNSKIM ELLQVVPSCV

```

901 ASLDDVAETD RKEYVTVKIR KWNNSKLYDL PDEPFTROFY FVHSAGQFKG  
951 KTSREIMARD RSVPNLTEGS LHEPGQSVT LRQNNLPAQS GSHAAEKGNS  
1001 DWPERPGLTQ TGPGRHMLR RHTVEDAVVS QGPEAAGPLS TPQEVASARS

## BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphtes3\_1817, frame 2

TREMBL:HSU43965.1 gene: "ANK3"; product: "ankyrin G119"; Human ankyrin  
G119 (ANK3) mRNA, complete cds., N = 2, Score = 287, P = 3.7e-21

PIR:I49502 ankyrin - mouse, N = 3, Score = 365, P = 2.2e-27

TREMBL:HSANKY\_2 product: "alt. ankyrin (variant 2.2)"; Human mRNA for  
ankyrin (variant 2.1), N = 2, Score = 380, P = 7.3e-31

SWISSPROT:ANK1\_HUMAN ANKYRIN R (ANKYRINS 2.1 AND 2.2) (ERYTHROCYTE  
ANKYRIN), N = 2, Score = 380, P = 8.2e-31

PIR:SJHUK ankyrin 1, erythrocyte splice form 1 - human, N = 2, Score =  
380, P = 8.2e-31

>TREMBL:HSANKY\_2 product: "alt. ankyrin (variant 2.2)"; Human mRNA for  
ankyrin (variant 2.1)  
Length = 1,719

## HSPs:

Score = 380 (57.0 bits), Expect = 7.3e-31, Sum P(2) = 7.3e-31  
Identities = 139/447 (31%), Positives = 207/447 (46%)

Query: 462 RGHTPLHVAACVCGQASLIDLLVSKGAMVNATDYHGATPLHLACQKGYQSVTLLLLHYKAS 521  
+G+T LH+AA+ GQ ++ LV+ GA VNA G TPL++A Q+ + V LL A+  
Sbjct: 77 KNTALHIAALAGQDEVVRELNVYGANVNAQSQKGFPLYMAAQENHLEVVKFLENGAN 136

Query: 522 AEVQDNNGNTPLHLACTYGHEDCVKALVYVDVES-CRL----- 558  
V +G TPL+A GHE+ V L+ Y + RL  
Sbjct: 137 QNVATEDGFTPLAVALQQGHENVVAHLINYGTKGVRLPALHIAARNDDTRTAAVLLQND 196

Query: 559 ---DIGNEKGDTPHLHIAARWGYQGVETLLONGASTEIQNRLKETPLKCALNSKILSVME 615  
D+ ++ G TPLHIAA + V + LL GAS + TPL A S+ +V+  
Sbjct: 197 PNPDLVSKTGFTPLHIAAHYENLNVAQLLLNRGASVNFTPNQGITPLHIA--SRRGNVIM 254

Query: 616 AYHLSFERRQKSSEAPVQSPQRSVDSISQESSTS-SFSSMSAGSR-QEETKKDYREVEKL 673  
L +R + E + + ++ S + G+ Q +TK +  
Sbjct: 255 V-RLLDLGAQI-ETKTKDELTPHCAARNGHVRISEILLDHGAPIQAKTKNGLSPIHM- 311

Query: 674 LRAVADGD-LEMVRYLLEWTEEDLEDAEDTVSAADPEFCHPLCQCPKCAPAQKRLAKVPA 732  
A GD L+ VR LL++ E ++D T+ P H C R+AKV  
Sbjct: 312 ---AAQGDHLDCVRLLLQYDAE-IDDI--TLDHLTP--LHVAHC-----GHHRVAKVLL 358

Query: 733 S-GLGVNVTSDQGSPLHVAALHGRADLIRLLKHGANAGARNADQAVPLHLACQGHFQ 791  
G N + +G +PLH+A ++ LLLK GA+ A PLH+A GH  
Sbjct: 359 DKGAKPNSRALNGFTPLHIACKKNHVRVMEILLKTGASIDAVTESGLTPLHVASFMGHL 418

Query: 792 VVKCLDSNAKPNKKDLSGNTPLIYACSGGHHVALLLQHGASINASNKNGNTALHEAV 851  
+VK LL A PN ++ TPL A GH E+ LLQ+ A +NA T LH A  
Sbjct: 419 IVKNLLQRGASPNVSNVKVETPLHMAARAGHTEVAKYLLQNKAKVNAKAKDDQTPHCAA 478

Query: 852 IEKHVFVVELLLHGASVQVNLNRQRTAVDCAEQSKINELLQVV 896  
H +V+LLL +A+ + T + A + + +L ++  
Sbjct: 479 RIGHTNMVRLLENNANPNLATTAGHTPLHIAAREGHVETVLALL 523

Score = 378 (56.7 bits), Expect = 1.2e-30, Sum P(2) = 1.2e-30  
Identities = 130/447 (29%), Positives = 195/447 (43%)

Query: 465 TPLHVAACVCGQASLIDLLVSKGAMVNATDYHGATPLHLACQKGYQSVTLLLLHYKASAEV 524  
TPLH AA G + ++L+ GA + A +G +P+H+A Q + LLL Y A +  
Sbjct: 274 TPLHCAARNGHVRISEILLDHGAPIQAKTKNGLSPIHMAAQDHLDCVRLLLQYDAEIDD 333

Query: 525 QDNNGNTPLHLACTYGHEDCVKALVYVDVE-----SCR----- 557  
+ TPLH+A GH K L+ + +C+  
Sbjct: 334 ITLDHLTPHVAACHGHRVAKVLLDQKAPNSRALNGFTPLHIACKKNHVRVMEILLKT 393

Query: 558 ---LDIGNEKGDTPHLHIAARWGYQGVETLLONGASTEIQNRLKETPLKCALNSKILSV 614  
+D E G TPLH+A+ G+ ++ LLQ GAS + N ETPL A + V  
Sbjct: 394 GASIDAVTESGLTPLHVASFMGHLPIVKNLLQRGASPNVSNVKVETPLHMAARAGHTEVA 453

Query: 615 EAYHLSFERRQKSSEAPVQSPORSVDISIQESSTSSFSMSAGSRQEETKKDYREVEKLL 674  
+ Y L + + + Q+P I + +A T L  
Sbjct: 454 K-YLLQNKAKVNAKAKDDQTPHCAARIGHTNMVKLLLENNANPNLATTAGH----TFLH 508

Query: 675 RAVADGDEMVRYLEWTEEDLEDAEDTSAADPEFCHPLCQCPKCAPAQKRLAKVPASG 734  
A +G +E V LLE ++ A T P H + K A+ L +  
Sbjct: 509 IAREGHVETVLALLE---KEASQACMTKKGFTP--LHVAARYGKVRVAELLER----D 559

Query: 735 LGVNVTSQDGSPLHVAALHGRADLIRLLKKGANAGARNADQAVPLHLACQGHFQVVK 794  
N ++G +PLHVA H D+++LLL G + + PLH+A +Q +V +  
Sbjct: 560 AHFNAAGKNGLTPLHVAHHNLDIVKLLPRGGSHPSPAWNGYTPHIAAKQNQVEVAR 619

Query: 795 CLLDSNAKPNKKDLSGNTPLIYACSGGHHHELVALLLQH GASINASNKGN TALHEAVIEK 854  
LL N + + G TPL A GH E+VALL A+ N N G T L H E  
Sbjct: 620 SLLQYGGSANAESVQGVTPHLAAQEGHAEMVALLSKQANGNLGKSGLTPLHLVAQEG 679

Query: 855 HVFVVELLLH GASVQVLNKRQRTAVDCAEQ--NSKIMELL 893  
HV V ++L+ HG V + T + A N K+++ L  
Sbjct: 680 HVPVADVLKKGVMVDATTRMGYTPHVAHYGNIKLVKFL 720

Score = 367 (55.1 bits), Expect = 1.8e-29, Sum P(2) = 1.8e-29  
Identities = 131/489 (26%), Positives = 210/489 (42%)

Query: 404 HIAS--GNQKEVERLLSQEDHDKDTVQKMCHPL-CFCDCEKLVSGRLNDPSVVTFFSRD 460  
HIAS GN V LL + + + PL C + +S L D ++  
Sbjct: 244 HIASRRGNVIMVRLLLDRGAQIETKTDELTPHCAARNGHVRSEILLDHGAPIQ-AKT 302

Query: 461 DRGHTPLHVAACVCGQASLIDLLVSKGAMVNATDYHGATPLHLACQKGYQSVTLLLLHYKA 520  
G +P+H+AA + LL+ A ++ TPLH+A G+ V +LL A  
Sbjct: 303 KNGLSPIHMAAQGDHLDCVRLLLQYDAEIDDITLDHLTPLHVAACHGHHRVAKVLLDKGA 362

Query: 521 SAEVQDNNGNTPLHLACTYGHEDCVKALVYVDVESCRLDIGNEKGDTPHIAARWGYQGV 580  
+ NG TPLH+AC H ++L+ +D E G TPLH+A+ G+ +  
Sbjct: 363 KPNRSLNGFTPLHIACKKNHVRVMELLK---TGASIDAVTESGLTPLHVASFMHGLPI 419

Query: 581 IETLLQNGASTEIQNRLKETPLKCAL--NSKILSVMEAYHLSFERRQKSSEAPVQSPQR 637  
++ LLQ GAS + N ETPL A ++++ + K + P+ R  
Sbjct: 420 VKNLLQRGASPNVSNVKVETPLHMAARAGHTEVAKYLLQNKAKVNAKAKDDQTPHCAAR 479

Query: 638 ---SVDSISQESSTSSFSMSAGSRQEETKKDYREVEKLLRAVADGDEMVRYLEWTE 693  
++ + E++ + + +AG VE +L + + +T  
Sbjct: 480 IGHNTNMVKLLLENNANPNLATTAGHTPLHIAAREGHVETVLALLEKEASQACMTKKGFTP 539

Query: 694 EDLEDAEDTSAAD---PEFCHPLCQ----CP-KCAPAQKRLAKVPA---SGLGVNVTS 741  
+ V A+ HP P A L V G + +  
Sbjct: 540 LHVAARYGKVRVAELLERDAHPNAAGKNGLTPLHVAHHNLDIVKLLPRGGSHPSPA 599

Query: 742 QDGSPLHVAALHGRADLIRLLKKGANAGARNADQAVPLHLACQGHFQVVKCLLDSNA 801  
+G +PLH+AA + ++ R LL++G +A A + PLHLA Q+GH ++V LL A  
Sbjct: 600 WNGYTPHIAAKQNQVEVARSLQYGGSANAESVQGVTPHLAAQEGHAEMVALLSKQA 659

Query: 802 KPNKKDLSGNTPLIYACSGGHHHELVALLLQH GASINASNKGN TALHEAVIEKHVVFVEL 861  
N + SG TPL GH + +L+HG ++A+ G T L H A ++ +V+  
Sbjct: 660 NGLGKSGSLTPLHLVAQEGHVPVADVLIKHGVMVDATTRMGYTPHVAHYGNIKLVKF 719

Query: 862 LLLH GASVQVLNK 874  
LL H A V K  
Sbjct: 720 LLQHQADVNAKTK 732

Score = 345 (51.8 bits), Expect = 4.2e-27, Sum P(2) = 4.2e-27  
Identities = 146/506 (28%), Positives = 233/506 (46%)

Query: 404 HIAS--GNQKEVERLLSQEDHDKDTVQK---MCHPLCFCDCEKLVSGRLNDPSVVTFFS 458  
H+AS G+ K V LL +E + T +K H +++V +N + V +  
Sbjct: 50 HLASKEGHVKMVVELLHKEIILETTTKGN TALHIALAQ-DEVRELNVYGANVN--A 106

Query: 459 RDRGHTPLHVAACVCGQASLIDLLVSKGAMVNATDYHGATPLHLACQKGYQSVTLLLLHY 518  
+ +G TPL+AA ++ L+ GA N G TPL +A Q+G+++V L++Y  
Sbjct: 107 QSQKGFPTPLMAAQENHLEVVKFLENGANQNVATEDGFTPLAVALQGHENVVHLIN 166

Query: 519 KASAEVQDNNGNTP-LHLACTYGHEDCVKALVYVDVESCRLDIGNEKGDTPHIAARWGY 577  
+V+ P LH+A ++D A V + D+ ++ G TPLHIA +  
Sbjct: 167 GTKGVR-----LPALHIAAR--NODTRTAVLLQNDP-NPDVLSKTGFTPLHIAAHYEN 218

Query: 578 QGVIELLLQNGASTEIQNRLKETPLKCAL--NSKILSVMEAYHLSFERRQKSSEAPVOS 634  
V + LL GAS + TPL A N ++ ++ E + K P+  
Sbjct: 219 LNVAQLLLNRGASVNFPTQNGITPLHIASRRGNVIMVRLLLDRGAQIETKTDELTPHLC 278

Query: 635 PQRSVDISIQESSTSSFSMSAGSRQEETKKDYREVEKLLRAVADGD-LEMVRYLEWTE 693  
R+ .E + + A +TK + A GD L+ VR LL++  
Sbjct: 279 AARNGHVRSEILLDHGAPIQA-----KTKNGLSPIHM-----AAQGDHLDCVRLLLQYDA 329

Query: 694 EDLEDAE-DTVSAAD-PEFC--HPLCQC-----PK-----CAPAQKRLAK 729  
E ++D D ++ C H + + P C R + +  
Sbjct: 330 E-IDDTLHDLTFLHVAHCGHHRVAKVLLDKGAKPNSRALNGFTPLHIACKKNHVRVME 388

Query: 730 VPA-SGLGVNVTSDGSSPLHVAALHGRADLIRLLKKGANAGARNADQAVPLHLACQCG 788  
+ +G ++ ++ G +PLHVA+ G +++ LL+ GA+ N PLH+A + G  
Sbjct: 389 LLLKTGASIDAVTESGLTPLHVASFMGHLPIVKNLLQRGASPNVSNVKEVETPLHMAARAG 448

Query: 789 HFQVVKCLLDSNAKPNKDLGNTPLIYACSGGHHELVALLQHGASINASNKGNTALH 848  
H +V K LL + AK N K TPL A GH +V LLL++ A+ N + G+T LH  
Sbjct: 449 HTEVAKYLLQNKAKVNAKAKDDQTPLHCAARIGHTNMVKLLLENNANPNLATTAGTAPLH 508

Query: 849 EAVIEKHVFVVELLLHGASVQVLNKRQRTAVDCAEQNSKIM--ELL 893  
A E HV V LL AS + K+ T + A + K+ ELL  
Sbjct: 509 IAAAREGHVETVLALLEKEASQACMTKKGFTPLHVAKYGKVRVAELL 555

Score = 243 (36.5 bits), Expect = 1.6e-14, Sum P(2) = 1.6e-14  
Identities = 64/199 (32%), Positives = 97/199 (48%)

Query: 404 HIAS--GNQKEVERLLSQEDHDKDTVQKMCHPLCFDDCEKLVSGRLNDPSVVTFFSRDD 461  
H+A+ G + E LL ++ H + PL L + L P + P S  
Sbjct: 541 HVAAYGKVRVAELLERDAHPNAAGKNGLTPLHVAVHNNLDIVKLLPRGGSPPHSPAW 600

Query: 462 RGHTPLHVAAVCGQASLIDLLVSKGAMVNATDYHGATPLHLACQKGYQSVTLLLLHYKAS 521  
G+TPLH+AA Q + L+ G NA G TPLHLA Q+G+ + LLL +A+  
Sbjct: 601 NYTTPHIAAKQNOVEVARSLQYGGSSANAESVQGVTPHLHAAQEGHAEMVALLSKQAN 660

Query: 522 AEVQDNNNGTPLHLACTYGHEDCVKALVYYDVESCRLDIGNEKGDTPHLIAARMWGYQGI 581  
+ + +G TPLHL GH L+ + V +D G TPLH+A+ +G ++  
Sbjct: 661 GNLGNKSGLTPLHLVAQEGHVPVADVLIKGV--MVDATTRMGYTPLHVASHYGNIKLV 717

Query: 582 ETLLQNGASTEIQNRKLTPL 602  
+ LLQ+ A + +L +PL  
Sbjct: 718 KFLQHQAQDVNAKTLGYSPL 738

Score = 242 (36.3 bits), Expect = 5.0e-29, Sum P(2) = 5.0e-29  
Identities = 63/176 (35%), Positives = 92/176 (52%)

Query: 734 GLGVNVTSDGSSPLHVAALHGRADLIRLLKKGANAGARNADQAVPLHLACQCGHFQV 793  
G VN T Q+G +PLH+A+ G ++RLLL GA + D+ PLH A + GH ++  
Sbjct: 229 GASVNFPTQNGITPLHIASRRGNVIMVRLLLDRGAQIETKTKDELTPHLCAARNGHVRIS 288

Query: 794 KCLLDSNAKPNKDLGNTPLIYACSGGHHELVALLQHGASINASNKGNTALHEAVIE 853  
+ LLD A K +G +P+ A G H + V LLQ+ A I+ T LH A  
Sbjct: 289 EILLDHGAPIQAKTKNGLSPIHMAAQGDHLDVRLLLQYDAEIDDTLHDLTFLHVAHC 348

Query: 854 KHVFVVELLLHGA--SVQVLNKRQRTAVDCAEQNSKIMELLQVVPSCVASLDDVAET 909  
H V ++LL GA + + LN + C + + +MELL AS+D V E+  
Sbjct: 349 GHHRVAKVLLDKGAKPNSRALNGFTPLHIACKKNHVRVMEMLLKTG---ASIDAVTES 403

Score = 242 (36.3 bits), Expect = 3.3e-14, Sum P(2) = 3.3e-14  
Identities = 80/284 (28%), Positives = 129/284 (45%)

Query: 404 HIAS--GNQKEVERLLSQEDHDKDTVQKMCHPLCFDDCEKLVSGRLNDPSVVTFFSRDD 461  
HIA+ G+ + V LL +E +K PL K+ L P +  
Sbjct: 508 HIAAREGHVETVLALLEKEASQACMTKKGFTPLHVAKYGKVRVAELLERDAHPNAAGK 567

Query: 462 RGHTPLHVAAVCGQASLIDLLVSKGAMVNATDYHGATPLHLACQKGYQSVTLLLLHYKAS 521  
G TPLHVA ++ LL+ +G ++ ++G TPLH+A ++ V LL Y S  
Sbjct: 568 NGLTPLHVAVHNNLDIVKLLPRGGSPPHSPAWNGYTPLHIAAKQNOVEVARSLQYGG 627

Query: 522 AEVQDNNNGTPLHLACTYGHEDCVKALVYYDVESCRLDIGNEKGDTPHLIAARMWGYQGI 581  
A + G TPLHLA GH + V L+ ++GN+ G TPLH+ A+ G+ V  
Sbjct: 628 ANAESVQGVTPHLHAAQEGHAEMVALLSKQANG---NLGNKSGLTPLHLVAQEGHVPVA 684

Query: 582 ETLLQNGASTEIQNRKLTPLKCAL---NSKILSVMEAYHLSFERRQKSEAPV-QSPQR 637  
+ L++G + R+ TPL A N K++ + + K +P+ Q+ Q+  
Sbjct: 685 DVLIKHGVMVDATTRMGYTPLHVASHYGNIKLVKFLQHQAQDVNAKTLGYSPLHQAQQ 744

Query: 638 S-VDSISQ--ESSTSFSSMSAGSQRQETTK--DYREVEKLLRAVAD 679  
D ++ ++ S S G+ K Y V +L+ V D  
Sbjct: 745 GHTDIVTLLKNGASPNEVSSDGTPLAIARLGYISVTDVLKVVTD 791

Score = 235 (35.3 bits), Expect = 7.9e-34, Sum P(2) = 7.9e-34  
Identities = 58/165 (35%), Positives = 83/165 (50%)

Query: 734 GLGVNVTSDGSSPLHVAALHGRADLIRLLKKGANAGARNADQAVPLHLACQCGHFQV 793  
G N S G +PLH+AA G A+++ LLL AN N PLHL Q+GH V  
Sbjct: 625 GGSANAESVQGVTPHLHAAQEGHAEMVALLSKQANGNLGNKSGLTPLHLVAQEGHVPVA 684

Query: 794 KCLLDSNAKPNKKDLSGNTPLIYACSGGHHELVALLLQH GASINASNKNGNTALHEAVIE 853  
L+ + G TPL A G+ +LV LLQH A +NA G + LH+A +  
Sbjct: 685 DVLIKHGVMVDATTRMGYTPLHVASHYGNIKLVKFL LQH QADVNAKTLGYSPLHQAAQ 744

Query: 854 KHVFVVELLLH GASVQVLNKRQRTAVDCAEQNS--KIMELLQVV 896  
H +V LLL +GAS ++ T + A++ + ++L+VV  
Sbjct: 745 GHTDIVTLLKNGASPNVSSDGTTP LAIAKRLGYISVTDVLKVV 789

Score = 233 (35.0 bits), Expect = 7.9e-34, Sum P(2) = 7.9e-34  
Identities = 67/202 (33%), Positives = 100/202 (49%)

Query: 404 HIAS-GNQKEVERLLSQEDHKDTVQKMCH--PLCFDDC-EKLVSGRLNDPSVVTFFSR 459  
H+A+ G+ + RLL Q D + D + + H PL C V+ L D P SR  
Sbjct: 310 HMAAQGDHLDVRLLLQYDAEIDDIT-LDHLTPHVAACHGHRVAKVLLDKGA-KPNSR 367

Query: 460 DDRGHTPLHVAACGQASLIDLLVSKGAMVNATDYGATPLHLACQKGYQSVTLLLLHYK 519  
G TPLH+A +++LL+ GA ++A G TPLH+A G+ + LL  
Sbjct: 368 ALNGFTPLHIACKKNHVRVMEILLKTGASIDAVTESGLTPLHVASFMGHLPIVKNLLQRG 427

Query: 520 ASAEVQDNNGNTPLHLACTYGHEDCVKALVYDVESCRLDIGNEKGDTPHIAARWGYQG 579  
AS V + TPLH+A GH + K L+ +++ + TPLH AAR G+  
Sbjct: 428 ASPNVSNVKTETPLHMAARAGHTEVAKYLLQ---NKAKVNAKAKDDQTPHCAARIGHTN 484

Query: 580 VIETLLQNGASTEIONRLKETPLKCA 605  
+++ LL+N A+ + TPL A  
Sbjct: 485 MVKLLLENNANPNLATTAGHTPLHIA 510

Score = 226 (33.9 bits), Expect = 7.0e-33, Sum P(2) = 7.0e-33  
Identities = 53/153 (34%), Positives = 83/153 (54%)

Query: 743 DGSSPLHVAALHGRADLIRLLKKGANAGARNADQAVPLHLACQGHFQVVKCLLDSNAK 802  
+G +PLH+AA + ++ R LL++G +A A + PLHLA Q+GH ++V LL A  
Sbjct: 601 NYGTPLHIAAKQNVQVEVARSLQYGSANAEVQGVTPHLAAQEGHAEMVALLSKQAN 660

Query: 803 PNKKDLSGNTPLIYACSGGHHELVALLLQH GASINASNKNGNTALHEAVIEKHVFVVELL 862  
N + SG TPL GH + +L++HG ++A+ G T LH A ++ +V+ L  
Sbjct: 661 GNLGKSGLTPLHLVAQEGHVPVADVLIKHGVMVDATTRMGYTPLHVASHYGNIKLVKFL 720

Query: 863 LLHGASVQVLNKRQRTAVDCAEQ--NSKIMELL 893  
L H A V K + + A Q ++ I+ LL  
Sbjct: 721 LQH QADVNAKTLGYSPLHQAAQGHDTIVTLL 753

Score = 198 (29.7 bits), Expect = 2.5e-11, Sum P(2) = 2.5e-11  
Identities = 51/157 (32%), Positives = 82/157 (52%)

Query: 737 VNVTSQDGSSPLHVAALHGRADLIRLLKKGANAGARNADQAVPLHLACQGHFQVVKCL 796  
+ T++ G++ LH+AAL G+ +++R L+ +GAN A++ PL++A Q+H +VVK L  
Sbjct: 71 LETTTKKGNTALHIAALAGQDEVVRELNVYGANVNAQSQKGTPLYMAAQENHLEVVKFL 130

Query: 797 LDSNAKPNKKDLSGNTPLIYACSGGHHELVALLLQH GASINASNKNGNTALHEAVIEKHV 856  
L++ A N G TPL A GH +VA L+ +G ALH A  
Sbjct: 131 LENGANQNVATEDGFTPLAVALQGGHENVVAHLINYGTK---GKVRPLPALHIAARNDDT 186

Query: 857 FVVELLLH GASVQVLNKRQRTAVDCAE--QNSKIMELL 893  
+LL + + VL+K + T + A +N + +LL  
Sbjct: 187 RTAAVLLQNDPNPDVLSKTGFTPLHIAAHYENLNVAQLL 225

Score = 186 (27.9 bits), Expect = 6.6e-29, Sum P(2) = 6.6e-29  
Identities = 55/143 (38%), Positives = 68/143 (47%)

Query: 463 GHTPLHVAACGQASLIDLLVSKGAMVNATDYGATPLHLACQKGYQSVTLLLLHYKASA 522  
GHTPLH+AA G + L+ K A G TPLH+A + G V LLL A  
Sbjct: 503 GHTPLHIAAREGHVETVLALLEKEASQACMTKKGFTPLHVAAYGKVRVAELLERDAHP 562

Query: 523 EVQDNNGNTPLHLACTYGHEDCVKALVYDVESCRLDIGNEKGDTPHIAARWGYQGVIE 582  
NG TPLH+A + + D VK L+ S N G TPLHIAA+ V  
Sbjct: 563 NAAGKNGLTPLHVAVHHNNLDIVKLLPRG-GSPHSPAWN--GYTPLHIAAKQNVQVEVAR 619

Query: 583 TLLQNGASTEIONRLKETPLKCA 605  
+LLQ G S ++ TPL A  
Sbjct: 620 SLLQYGSANAEVQGVTPHLHA 642

Score = 182 (27.3 bits), Expect = 2.9e-28, Sum P(2) = 2.9e-28  
Identities = 54/185 (29%), Positives = 89/185 (48%)

Query: 738 NVTSQDGSSPLHVAALHGRADLIRLLKKGANAGARNADQAVPLHLACQGHFQVVKCLL 797  
N+ ++ G +PLH+ A G + +L+KHG A PLH+A G+ ++VK LL  
Sbjct: 662 NLGNKSGLTPLHLVAQEGHVPVADVLIKHGVMVDATTRMGYTPLHVASHYGNIKLVKFL 721

Query: 798 DSNAPNKKDLSGNTPLIYACSGGHHELVALLLQH GASINASNKNGNTALHEAVIEKHVF 857  
A N K G +PL A GH ++V LLL++GAS N ++ G T L A ++

Sbjct: 722 QHQADVNAKTKLGYSPHQAQQGHTDIVTLLKNGASPNVSSDGTTPLAIAKRLGYIS 781

Query: 858 VVELLLHGGASVQLNKRQRTAVDCAEQNSKIMELLQVVPSCVASLDDVAETDRKEYVTV 917  
V ++L + V ++ V + S P V + DV+E + +E ++

Sbjct: 782 VTDVLKV-----VTDETSFVLVSDKHRMS-----FPETVDEILDVSEDEGEELISF 827

Query: 918 KIRKK 922  
K ++

Sbjct: 828 KAERR 832

Score = 180 (27.0 bits), Expect = 5.0e-29, Sum P(2) = 5.0e-29  
Identities = 41/121 (33%), Positives = 67/121 (55%)

Query: 486 GAMVNATDYHGATPLHLACQKGYQSVTLLLLHYKASAEVDNNGNTPLHLACTYGHEDCV 545  
G +N + +G LHLA ++G+ + + LLH + E GNT LH+A G ++ V

Sbjct: 35 GVDINTCNQNLGLHLASKEGHVKMVVELLHKEIILETTTCKGNTALHIAALAGQDEVV 94

Query: 546 KALVYYDVESCRLDIGNEKGTPLHIAARWGYQGVETLLONGASTEIQNRLKETPLKCA 605  
+ LV Y ++ ++KG TPL+AA+ + V++ LL+NGA+ + TPL A

Sbjct: 95 RELVNY---GANVNAQSQKGFPLYMAAQENHLEVVKFLLENGANQNVATEDGFTPLAVA 151

Query: 606 L 606  
L

Sbjct: 152 L 152

Score = 166 (24.9 bits), Expect = 3.4e-06, Sum P(2) = 3.4e-06  
Identities = 89/318 (27%), Positives = 140/318 (44%)

Query: 448 LNDPSVVTFFSRDDRGHTPLHVAACGQASLIDLLVSKGAMVNATDYHGATPLHLACQK 507  
L + + V ++DD+ TPLH AA G +++ LL+ A N G TPLH+A ++G

Sbjct: 457 LQNKAKVNAKAKDDQ--TPLHCAARIGHTNMVKLLLENNANPNLATTAGHTPLHIAAREG 514

Query: 508 YQSVTLLLLHYKASAEVDNNGNTPLHLACTYGHEDCVKALVYYD----- 552  
+ L LL +AS G TPLH+A YG + L+ D

Sbjct: 515 HVETVLALKEASQACMTKKGFTPLHVAAYGKVRVAELLERDAHPNAAGKGLTPLH 574

Query: 553 --VESCRLDI-----GNE-----KGTPLHIAARWGYQGVETLLONGASTEIQNRL 597  
V LDI G+ G TPLHIAA+ V +LLQ G S ++

Sbjct: 575 VAVHHNNLDIVKLLPRGSPHSPAWNCTPLHIAAKQNOVEVARSLQYGGSAEAESVQ 634

Query: 598 KETPLKCALNSKILSVMEAYHLSFERROKSSAPVQSPQRSVDSISQESSTSSFSM-SA 656  
TPL A +M A LS +Q + +S + ++QE +

Sbjct: 635 GVTPLHLAAQEGHAE-MVALLS---KQANGNLGKSGLTPLHLVAQEGHVPVADVLIKH 690

Query: 657 GSRQEETKKDYREVEKLLRAVADGDLEVMRYLLEWTEEDLEDAEDTVAADPEFCHPLCO 716  
G + T + L A G++++V++LL+ + D+ +A+ + + PL Q

Sbjct: 691 GVMVDATTR--MGYTPLVASHYGNIKLVKFLQH-QADV-NAKTKLGS-----PLHQ 740

Query: 717 CPKCAPAQKRLAKVPASGLGVNVTSQDGSSPLHVA 751  
+ + +G N S DG++PL +A

Sbjct: 741 AAQQGHTDI-VTLLKNGASPNVSSDGTTPLAIA 774

Score = 162 (24.3 bits), Expect = 1.8e-07, Sum P(2) = 1.8e-07  
Identities = 48/149 (32%), Positives = 71/149 (47%)

Query: 737 VNVTSQDGSSPLHVAALHGRAOLIRLLKXGAGARNADQAVPLHLACQGHFQVVKCL 796  
V D ++ AA G D L++G + N+ LHLA ++GH ++V L

Sbjct: 5 VGFREDAATSFLLAARSGNLDKALDHLRNGVDINTCNQNLGLHLASKEGHVKMVVEL 64

Query: 797 LDSNAKPNKKDLGNTPLIYACSGGHELVALLQHGASINASNKNTALHEAVIEKHV 856  
L GNT L A G E+V L+ +GA++NA + KG T L+ A E H+

Sbjct: 65 LHKEIILETTTCKGNTALHIAALAGQDEVVRELVNYGANVNAQSQKGFPLYMAAQENHL 124

Query: 857 FVVELLLHGGASVQLNKRQRTAVDCAEQ 885  
VV+ LL +GA+ V + T + A Q

Sbjct: 125 EVVKFLLENGANQNVATEDGFTPLAVALQ 153

Score = 158 (23.7 bits), Expect = 5.7e-26, Sum P(2) = 5.7e-26  
Identities = 38/135 (28%), Positives = 65/135 (48%)

Query: 460 DDRGHTPLHVAACGQASLIDLLVSKGAMVNATDYHGATPLHLACQKGYQSVTLLLLHYK 519  
+ G LH+A+ G ++ L+ K ++ T G T LH+A G V L++Y

Sbjct: 42 NQNLGLHLASKEGHVKMVVELLHKEIILETTTCKGNTALHIAALAGQDEVVRELVNYG 101

Query: 520 ASAEVDNNGNTPLHLACTYGHEDCVKALVYYDVESCRLDIGNEKGTPLHIAARWGYQG 579  
A+ Q G TPL+AA H + VK L+ ++ E G TPL +A + G++

Sbjct: 102 ANVNAQSQKGFPLYMAAQENHLEVVKFLE---NGANQNVATEDGFTPLAVALQOQHEN 158

Query: 580 VIETLLONGASTEIQ 594  
V+ L+ G +++

Sbjct: 159 VVAHLINYGTKGKVR 173

Score = 115 (17.3 bits), Expect = 1.8e-21, Sum P(2) = 1.8e-21  
Identities = 37/119 (31%), Positives = 58/119 (48%)

Query: 497 ATPLHLACQKGYQSVTLILLHYKASAEVQ--DNNNGTPLHLACTYGHEDCVKALVYYDVE 554  
AT A + G ++ L H + ++ + NG LHLA GH V L++ ++  
Sbjct: 13 ATSFLLRAARSG--NLDKALDHLRNGVDINTCNQNLGLHLASKEGHVKMVELLHKEII 70

Query: 555 SCRLDIGNEKGDTPHIAARWGYQGVETLLQNGASTEIQNRKETPLKCALNSKILSV 614  
L+ +KG+T LHIAA G V+ L+ GA+ Q++ TPL A L V+  
Sbjct: 71 ---LETTTKGNTALHIAALAGQDEVVRELNVYGANVNAQSQKGFPLYMAAQENHLEV 127

Query: 615 E 615  
+  
Sbjct: 128 K 128

Score = 106 (15.9 bits), Expect = 1.8e-01, Sum P(2) = 1.6e-01  
Identities = 34/121 (28%), Positives = 54/121 (44%)

Query: 769 NAGARNADQAVPLHLACQGHFQVVKLLDSNAKPNKKDLSCNTPLIYACSGGHHLEVAL 828  
+ G R AD A A + G+ L + N + +G L A GH ++V  
Sbjct: 4 SVGFREADAATSFLRAARSGNLDKALDHLRNGVDINTCNQNLGLHLASKEGHVKMVE 63

Query: 829 LLQHGASINASNKGNLALHEAVIEKHVVFVVELLLHGASVQLNKRQRTAVDCAEQNSK 888  
LL + + KGNTALH A + VV L+ +GA+V +++ T + A Q +  
Sbjct: 64 LLHKEIILETTTKGNTALHIAALAGQDEVVRELNVYGANVNAQSQKGFPLYMAAQENH 123

Query: 889 I 889  
+  
Sbjct: 124 L 124

Score = 40 (6.0 bits), Expect = 1.6e-14, Sum P(2) = 1.6e-14  
Identities = 11/56 (19%), Positives = 23/56 (41%)

Query: 622 ERRQKSSEAPVQSPQRSVDSISQESSTSSFSMSAGSRQEEKKDYREVEKLLRAV 677  
+RRQ+ E VQ + + + Q + + + Q ++ +K++R V  
Sbjct: 1614 DRRQQGQEQVQEAQNTFTQVVGNEFQNPGEQVTEEQFTDEQGNIVTKKIIRKV 1669

Score = 38 (5.7 bits), Expect = 2.6e-14, Sum P(2) = 2.6e-14  
Identities = 6/12 (50%), Positives = 10/12 (83%)

Query: 806 KDLSGNTPLIYA 817  
+D++G T L+YA  
Sbjct: 1186 EDITGTTLVYA 1197

Pedant information for DKFZphtes3\_1817, frame 2

#### Report for DKFZphtes3\_1817.2

[LENGTH] 1050  
[MW] 117013.72  
[PI] 6.47  
[HOMOL] TREMBL:DMANKY\_1 product: "ankyrin"; Drosophila melanogaster ankyrin mRNA,  
complete cds. 2e-45  
[FUNCAT] 08.19 cellular import [S. cerevisiae, YOR034c] 5e-13  
[FUNCAT] 10.05.99 other pheromone response activities [S. cerevisiae, YDR264c]  
3e-12  
[FUNCAT] 03.07 pheromone response, mating-type determination, sex-specific proteins  
[S. cerevisiae, YDR264c] 3e-12  
[FUNCAT] 99 unclassified proteins [S. cerevisiae, YIL112w] 2e-11  
[FUNCAT] 06.13.01 cytoplasmic degradation [S. cerevisiae, YGR232w] 8e-10  
[FUNCAT] 30.10 nuclear organization [S. cerevisiae, YIR033w] 2e-08  
[FUNCAT] 04.05.01.07 chromatin modification [S. cerevisiae, YIR033w] 2e-08  
[FUNCAT] 01.04.04 regulation of phosphate utilization [S. cerevisiae, YGR233c]  
3e-08  
[FUNCAT] 08.13 vacuolar transport [S. cerevisiae, YML097c] 5e-05  
[FUNCAT] 06.04 protein targeting, sorting and translocation [S. cerevisiae, YML097c]  
5e-05  
[FUNCAT] 30.03 organization of cytoplasm [S. cerevisiae, YML097c] 5e-05  
[FUNCAT] 08.07 vesicular transport (golgi network, etc.) [S. cerevisiae, YML097c]  
5e-05  
[FUNCAT] 03.22 cell cycle control and mitosis [S. cerevisiae, YER111c] 3e-04  
[FUNCAT] 04.05.01.04 transcriptional control [S. cerevisiae, YER111c] 3e-04  
[BLOCKS] BL00901A Cysteine synthase/cystathionine beta-synthase P-phosphate att  
dlawcb 1.91.3.1.2 GA binding protein (GABP) alpha GA bindini 4e-12  
[SCOP] 3.1.3.53 Myosin-light-chain-phosphatase 1e-12  
[EC] phosphotransferase 1e-19  
[PIRKW] nucleus 1e-13

[PIRKW] potassium channel 5e-15  
 [PIRKW] early protein 2e-13  
 [PIRKW] tumor suppressor 1e-09  
 [PIRKW] duplication 1e-14  
 [PIRKW] tandem repeat 1e-19  
 [PIRKW] heterodimer 1e-14  
 [PIRKW] potassium transport 5e-15  
 [PIRKW] cell cycle control 1e-10  
 [PIRKW] serine/threonine-specific protein kinase 1e-19  
 [PIRKW] transmembrane protein 5e-15  
 [PIRKW] transport protein 5e-15  
 [PIRKW] DNA binding 2e-11  
 [PIRKW] oncogene 1e-08  
 [PIRKW] ATP 1e-19  
 [PIRKW] protein kinase inhibitor 1e-09  
 [PIRKW] voltage-gated ion channel 5e-15  
 [PIRKW] phosphoprotein 4e-38  
 [PIRKW] apoptosis 1e-19  
 [PIRKW] liver 4e-09  
 [PIRKW] integrin binding 3e-16  
 [PIRKW] differentiation 2e-12  
 [PIRKW] transforming protein 1e-08  
 [PIRKW] alternative splicing 1e-40  
 [PIRKW] coiled coil 1e-14  
 [PIRKW] peripheral membrane protein 2e-38  
 [PIRKW] transcription factor 4e-16  
 [PIRKW] transcription regulation 2e-16  
 [PIRKW] nucleotide binding 5e-15  
 [PIRKW] phosphoric monoester hydrolase 1e-12  
 [PIRKW] cytoskeleton 8e-39  
 [PIRKW] calmodulin binding 1e-19  
 [PIRKW] smooth muscle 1e-12  
 [SUPFAM] ankyrin 1e-40  
 [SUPFAM] death-associated protein kinase 1e-19  
 [SUPFAM] ankyrin repeat homology 1e-40  
 [SUPFAM] protein kinase homology 1e-19  
 [SUPFAM] vaccinia virus 27.4K HindIII-C protein homology 3e-07  
 [SUPFAM] int-3 transforming protein 1e-08  
 [SUPFAM] unassigned ankyrin repeat proteins 2e-38  
 [SUPFAM] notch protein 2e-12  
 [SUPFAM] fowlpox virus BamHI-ORF7 protein 2e-13  
 [SUPFAM] rel homology 2e-11  
 [SUPFAM] EGF homology 2e-12  
 [PROSITE] ATP\_GTP\_A 1  
 [PFAM] Ank repeat  
 [KW] Irregular  
 [KW] 3D  
 [KW] LOW\_COMPLEXITY 3.05 %

SEQ MALYDELLKNPFYALQKCRPDLCCKVAQIHGIVLPCKGSLSSSIQSTCFESYILIP  
 SEG .....  
 lawCB .....  
  
 SEQ VEEHFQTLNGKDVFIQGNRIKLGAGFACLLSVLPILFEETFYNEKEESFSLCIAHPLEKR  
 SEG .....  
 lawCB .....  
  
 SEQ ESSEEPLAPSDPFSLKTIEDVREFLGRHSEFDRNIASFHRTFRECERKSLRHHIDSANA  
 SEG .....  
 lawCB .....  
  
 SEQ LYTQCLQQLLRDLSHLKMLAKQEAQMNLMKQAVEIYVHHEIYNLI FKYYVTMEASEDAAFN  
 SEG .....  
 lawCB .....  
  
 SEQ KITRSLQDLQKDIGVKPEFSFNI PRAKRELAQLNKCTSPQOKLVCLRKVVQLITQSPSQ  
 SEG .....  
 lawCB .....  
  
 SEQ RVNLETWCADDLLSVLLYLLVKTEIPNMNANLSYIKNFRFSSSLAKDELGYCLTSFEAAIE  
 SEG .....  
 lawCB .....  
  
 SEQ YIRQGSLSAKPPESEGFQDLFLKQMSLLSQMTSSPTDCLFKHIASGNQEVERLLSQE  
 SEG .....  
 lawCB .....  
  
 SEQ DHDKDTVQKMCPLCFDDCEKLVSGRLNDPSVVTFFSRDDRGHTPLHVAACVCGQASLID  
 SEG .....  
 lawCB .....



```

SEQ      LLVSKGAMVNATDYHGATPLHLACQKGYQSVTLLLLHYKASAEVDNNGNTPLHLACTYG
SEG      .....
lawcB    .....

SEQ      HEDCVKALVYYDVESCRDLGNEKGDTPHIAARMGYQGVETLLQNGASTEIQNRLKET
SEG      .....
lawcB    .....

SEQ      PLKCALNSKILSVMEAYHLSFERRQKSSEAPVQSPQRSVDSISQESSTSSFSMSAGSRQ
SEG      .....XXXXXXXXXXXXXXXXXXXXX.....
lawcB    .....

SEQ      EETKKDYREVEKLLRAVADGDLEMVRYLLEWTEEDLEDAEDTVSAADPEFCHPLCQCPKC
SEG      .....
lawcB    .....

SEQ      APAQKRLAKVPASGLGVNVTSDQGSSPLHVAALHGRADLIRLLKKGANAGARNADQAVP
SEG      .....
lawcB    .....CHHHHHHHHHHHCHHHHHHHHHCCCC-CCTTTTCCH

SEQ      LHLACQGHFQVVKCLLDSNAKPNKKDLSGNTPLIYACSGGHELVALLQHGASINASN
SEG      .....
lawcB    HHHHHHHCHHHHHHHHHCHCCCTTTCTTTTCHHHHHHHHTTHHHHHHHCHCCCTTTTEE

SEQ      NKGNTALHEAVIEKHVFVVELLLHGASVQVLNKRQRTAVDCAEQNSKIMELLQVVPSCV
SEG      .....
lawcB    TTTEHHHHHHCHHHHHHHHHCHCCCTTTTCBTTTCHHHHHHHCHCHHHHC.....

SEQ      ASLDDVAETDRKEYVTVKIRKWNKLYDLDEPFTRQFYFVHSAGQFKGKTSREIMARD
SEG      .....
lawcB    .....

SEQ      RSVPNLTGSLHEPGRQSVTLRQNNLPAQSGSHAAEKGNSDWPERPGLTQTGPGHRRMLR
SEG      .....
lawcB    .....

SEQ      RHTVEDAVSQGPEAAGPLSTPQEVASRS
SEG      .....
lawcB    .....

```

## Prosites for DKFZphtes3\_1817.2

PS00017      945->953      ATP\_GTP\_A      PDOC00017

## Pfam for DKFZphtes3\_1817.2

```

HMM_NAME      Ank repeat
HMM            *GyTPLHIAARYNNVEMVrLLQHGADIN*
               G+TPLH+AA ++ +++LL++GA +N
Query         463 GHTPLHVAAVCGQASLIDLLVSKGAMVN      490

32.12 (bits) f: 496 t: 523 Target: dkfzphes3_1817.2 similarity to ankyrins
Alignment to HMM consensus:
Query         *GyTPLHIAARYNNVEMVrLLQHGADIN*
               G TPLH+A++ + ++ LLL + A+
dkfzphes3     496 GATPLHLACQKGYQSVTLLLLHYKASAE      523

Query         f: 529 t: 556 Target: dkfzphes3_1817.2 similarity to ankyrins
Alignment to HMM consensus:
HMM           *GyTPLHIAARYNNVEMVrLLQHGADIN*
               G+TPLH+A+ Y+++++V+ L+ +
Query         529 GNTPLHLACTYGHEDCVKALVYYDVESC      556

42.65 (bits) f: 565 t: 592 Target: dkfzphes3_1817.2 similarity to ankyrins
Alignment to HMM consensus:
Query         *GyTPLHIAARYNNVEMVrLLQHGADIN*
               G+TPLHIAAR + +++ LLQ+GA+
dkfzphes3     565 GDTPLHIAARMGYQGVETLLQNGASTE      592

Query         f: 744 t: 771 Target: dkfzphes3_1817.2 similarity to ankyrins
Alignment to HMM consensus:
HMM           *GyTPLHIAARYNNVEMVrLLQHGADIN*
               G +PLH+AA +++ +++RLL+HGA+
Query         744 GSSPLHVAALHGRADLIRLLKKGANAG      771

```

36.38 (bits) f: 777 t: 804 Target: dkfzphtes3\_1817.2 similarity to ankyrins  
Alignment to HMM consensus:  
Query \*GyTPLHIAARYNNvEMVrILLQHGADIN\*  
PLH+A++++ ++V+ LL+ +A +N  
dkfzphtes3 777 QAVPLHLACQGGHfQVVKCLLDSNAKPN 804

Query f: 810 t: 837 Target: dkfzphtes3\_1817.2 similarity to ankyrins  
Alignment to HMM consensus:  
HMM \*GyTPLHIAARYNNvEMVrILLQHGADIN\*  
G+TPL++A+ ++ E+V LLLQHG+IN  
Query 810 GNTPLIYACSGGHHELVALLLQHGA SIN 837

44.62 (bits) f: 843 t: 870 Target: dkfzphtes3\_1817.2 similarity to ankyrins  
Alignment to HMM consensus:  
Query \*GyTPLHIAARYNNvEMVrILLQHGADIN\*  
G+T+LH A+++ +V +V+LLL HGA++  
dkfzphtes3 843 GNTALHEAVIEKHVFVVELLLHGASVQ 870

DKFZphtes3\_19f19  
-----

group: testes derived

DKFZphtes3\_19f19 encodes a novel 254 amino acid protein with weak similarity to *S. cerevisiae* protein YFL046w.

The protein contains a RGD cell attachment site.  
No informative BLAST results; No predictive prosite, pfam or SCOP motive.

The new protein can find application in studying the expression profile of testis-specific genes.

similarity to YFL046w

localisation: 3 STS match perfect but HS1292427 matches to chromosome 4

Sequenced by MediGenomix

Locus: /map="405.0/.3 cR from top of Chr11 linkage group"

Insert length: 1395 bp  
Poly A stretch at pos. 1367, no polyadenylation signal found

```

1 GGGACCACGG TGGCGCCTGC GCTGGGAGGT GAGCTTGTGA CAGAGCGAAA
51 ACTACAATTC CCAGCATTCC TGTGGTGCCA GAACCTACCTT GCCCGAAAAGC
101 CTGTGCGGAGA TTTACCCCGT CTTCGCCCTC CCTCCACCG GAAAACTCTG
151 AGGACATGAA TAGTCGCCAG GCTTGGCGGC TCTTCTCTC CCAAGGCAGA
201 GGAGATCGTT GGGTTTCAAG GCCCGCGGG CATTCTCGC CGGCCCTGGG
251 GAGAGAGTTC TTCACTACCA CAACCAAGGA GGGATATGAT AGGCGGCCAG
301 TGGATATAAC TCCTTTAGAA CAAAGGAAT TAACCTTTGA TACCCATGCA
351 TTGGTTGAGG ACTTGGAAAC TCATGGATT GACAAAACAC AAGCAGAAAC
401 AATTGTATCA GCGTTAACTG CTTTATCAAA TGTCAGCCTG GATACATCT
451 ATAAAGAGAT GGTCACTCAA GCTCAACAG AAATAACAGT ACAACAGCTA
501 ATGGCTCATT TGGATGCTAT CAGGAAAGAC ATGGTCATCC TAGAGAAAG
551 TGAATTTGCA AATCTGAGAG CAGAGAATGA GAAATGAAA ATTGAATTAG
601 ACCAAGTTAA GCAACAACCTA ATGCATGAAA CCAGTCGAAT CAGAGCAGAT
651 AATAAAGTGG ATATCAACTT AGAAAGGAGC ACAGTAACAG ATATCTTTAC
701 AGATCAAGAA AAGCAACTTA TGGAAACAC TACAGAATTT ACAAAAAGG
751 ATACTCAAAC CAAAGCTATT ATTTAGAGA CCAGTAATTA AATTGACGCT
801 GAAATTGCTT CCTTAAACAC ACTGATGGAA TCTAACAAAC TTGAGACAAT
851 TCGTTATCTT GCAGCTTCGG TGTTTACTTG CCGCAATA GCATTGGGAT
901 TTTTAGATT CTGCAAGTAG TATTAAATGCT CATCTGCTG TGGCTGTTGG
951 CTTCTTAGAA CACCAACCG GGAGAGATT ACTTTGAACA TTGCTAGTTG
1001 CAGCAAAAT TTAACAACA AGATTATTCG AAGTGTATAC GGACTAAAAG
1051 AGGAAGTGTT TTAGAATGAG AAGAGATCT GTGTCTTAT TGTGTGTGTG
1101 TGAGTGCAGG TGTGTGCTT TATTATATTG AAAAGCTGTC ACTCAGACCT
1151 GGTTTGAGAT AGAAGAGCAT TTTGTCCTT TGATAGTTAA TAGAAATTGA
1201 ACCAGAGTTT TCTTATGTTT GCTTGAACAG TTGTGTAAT CATACAGGAT
1251 TTTGTGGGTA TTGGTTGAAT ATTTGTAAC CATTCCCTAG CCTACATATT
1301 TATTACTGAA TTAACCTTCC TGATAACCAT TGCATAATTA CATTTTCTTA
1351 TAAATGAAA GATTATTACA AAAAAA AAAA AAAA

```

# BLAST Results

Entry HS419346 from database EMBL:  
human STS WI-13569.  
Score = 2154, P = 8.6e-91, identities = 446/459

Entry HS1292427 from database EMBL:  
human STS SHGC-50338.  
Score = 1737, P = 7.2e-72, identities = 359/369

Entry HS253344 from database EMBL:  
human STS WI-13893.  
Score = 1578, P = 1.0e-64, identities = 358/397

# Medline entries

No Medline entry

100 200 300 400 500 600 700 800 900 1000 1100 1200 1300 1400 1500 1600 1700 1800 1900 2000 2100 2200 2300 2400 2500 2600 2700 2800 2900 3000 3100 3200 3300 3400 3500 3600 3700 3800 3900 4000 4100 4200 4300 4400 4500 4600 4700 4800 4900 5000 5100 5200 5300 5400 5500 5600 5700 5800 5900 6000 6100 6200 6300 6400 6500 6600 6700 6800 6900 7000 7100 7200 7300 7400 7500 7600 7700 7800 7900 8000 8100 8200 8300 8400 8500 8600 8700 8800 8900 9000 9100 9200 9300 9400 9500 9600 9700 9800 9900 10000

```

1  MNSRQAWRLF  LSQGRGDRWV  SPRGRHFSPA  LRREFFTTTT  KEGYDRRPVD
51  TPLEQRKLK  FDTHALVQOL  EFGDFKPTQA  EITVSALTAL  SNVSLDTIYK
101  EMVTAQOEI  TVQMLHLDL  ATRKDMVILE  KSEFANLRAE  NEKMIELDQO
151  VKQQLMHETS  RIRADNKLDI  NLERSRVTDM  FTDQEKQLME  TTTEFTKKDT
201  QTKSISETS  NKIDAEIASL  KTLMESNKLE  TIRYLAASVF  TCLALALGFI
251  RFYK

```

No BLASTP hits available

SWISSPROT:YAN8 SCHPO HYPOTHETICAL 24.6 KD PROTEIN C3H1.08 IN CHROMOSOME  
I., N = 1, Score = 144, P = 8.4e-09

>SWISSPROT:YAN8\_SCHPO HYPOTHETICAL 24.6 KD PROTEIN C3H1.08 IN CHROMOSOME 1.  
Length = 211

Score = 144 (21.6 bits), Expect = 8.4e-09, P = 8.4e-09  
Identities = 34/121 (28%), Positives = 67/121 (55%)

```

Query: 70 LETHGFKDQOETIIVSALTALNSVSLDITIKEMVTOAQOE-ITVQOLMAHLDAIRKDMVI 128
      1E +G AETI + + ++ +L +K + +A+OE ++ QO L IRK +
Sbjct: 46 LEQAGYSVKNAETIITINLMRTITGEALTLEKNI GFKAQKQESV FQKRTFLQ-IRKYLET 104

Query: 129 LEKSEFANLRAENKMEKMIELDOVKQQLMHETSIRADNKLINDLERSVR TDMFTDQEKQL 188
      +E++EF +R ++K +E++ K L + ++ ++L+NL+R +D +T + +
Sbjct: 105 IEENEFDKVRKSSDKLNEIEKTKSSLREDVKTALSEVRLWNLNLEKGRMKDAATSRNTNI 164

Query: 189 ME 190
      E
Sbjct: 165 HE 166

```

.....

```

[LENGTH]      254
[MW]           29505.73
[PI]           6.99
[HOMOL]        PIR:S56209 probable membrane protein YFL046w - yeast (Saccharomyces cerevisiae)
2e-10
[FUNCAT]       99 unclassified proteins          [S. cerevisiae, YFL046w] 8e-12
[PROSITE]      RGD      1
[KW]            TRANSMEMBRANE 1
[KW]            LOW COMPLEXITY      5.12 %
[KW]            COILED COIL        11.02 %

```

```
SEQ      MNSRQAWRLFLSQSGDRWVSRPRGHFSPALRRFFTTTTTKEGYDRRPVDITPLEQRKLT
SEG
PRD      cccchhhhhhhhccccceeecccchhhhhhheeecccccccccfchhhhhc
COILS    .....
MEM      .....

SEQ      FDTHALVQDLETHGFDKTOAETIVSALTALSINVSLDTIYKEMVTQAQOEITVQQLMHAHL
SEG
PRD      chhhhhhhhhccccchhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh
COILS    .....
```

**PCT/IB00/01496**

Prosites for DKFZphtes3\_19f19.3

PS00016      15->18      RGD      PDOC00016

(No Pfam data available for DKFZphtes3\_19f19.3)

DKF2phtes3\_19j17  
-----

group: testes derived

DKF2phtes3\_19j17 encodes a novel 436 amino acid protein with partial similarity to C.elegans Y40B1A.2 protein.

The novel protein contains two Prosite WW/rsp5/WWP domain signatures. The WW domain (or rsp5 or WWP domain) has been originally discovered as a short conserved region in a number of unrelated proteins, such as dystrophin, utrophin, vertebrate YAP protein, mouse NEDD-4 and yeast RSP5. The domain is repeated up to 4 times in some proteins. It has been shown to bind proteins with particular proline-motifs, [AP]-P-P-[AP]-Y, and thus resembles somewhat SH3 domains. It appears to contain beta-strands grouped around four conserved aromatic positions; generally Trp. The name WW or WWP derives from the presence of these Trp as well as that of a conserved Pro. It is frequently associated with other domains typical for proteins in signal transduction processes.

No informative BLAST results; No predictive prosite, pfam or SCOP motif.

The new protein can find application in studying the expression profile of testis-specific genes.

similarity to C.elegans Y40B1A.2

there are two long ORFs in this cDNA according to EST:  
HS12146/HS75086/AA923755/MMAA17335 remaining intron at Bp 1506-1733

Sequenced by MediGenomix

Locus: unknown

Insert length: 2762 bp  
Poly A stretch at pos. 2740, no polyadenylation signal found

```
1 ATCTCAGCC AAATTTTTT ATTTTGGCA GAATCAGTG GCAAGTGGT
51 TTATAAGATA ATGGAGTGGT TTTTTTTGT GTTAGTGTG AITGTTATC
101 AGGAGTCTTA TTGTAACGGT TAAGCATTAG GTTTTTTGTG TGAGAAACTT
151 TAAAGAGTAA AGCAGAATTG AAAGTGGAAA TTTTAATTTT GTAAGTTCAT
201 AAAATTAAAT GATAATACAC CAAAGTTTAT GTTTAAATTA GGGAGTTTAA
251 GGTTCATTTT CTTCTCTTTT TTTTGGGGG GGGTCATGTT TTACAGGCAC
301 TTAAGTATTC ATCGAAGAGT CACCCAGTGA CCGGTGATCA CAGACATGAA
351 AAGATGCGAG ACGCCGGAGA TCCTTCACCA CCAATAAATA TGTTCGGGAG
401 ATCTGATAGT CCTGAAACAA AATACAGTGA CAGCACAGGT CACAGTAAGG
451 CCAAAATAGT GCATCTCAC AGAGTTAGAG AGAGGGATGG TGGGACCACT
501 TACTCTCCAC AAGAAATTC ACACAACCC AGTGCCTTTC ATAGTTCAAA
551 TTCACATTCT TCTAATCCAA GCAATAACCC AAGCAAACT TCAGATGCAC
601 CTTATGATTC TCCAGATGAC TGGTCTGAGC ATATTAGTTC TTCTGGGAAA
651 AAGTACTACT ACAATTGTCG AACAGAAGTT TCACAATGGG AAAAACCAAA
701 AGACTGGCTT GAAAGAGAAC AGAGACAAA AGAAGCAAGC AAGATGCCAG
751 TCAACAGCTT CCAAAAGAT AGGGATTACA GAAGAGAGGT GATGCAAGCA
801 ACAGCCACTA GTGGGTTTGC CAGTGGAAAG GAAGACAAGC ATTCCAGTGA
851 TGCCAGTAGT TTGCTCCAC AGAATATTTT GTCTCAAAAC AGCAGACACA
901 ATGACAGAGA CTACAGACTG CCAAGAGCAG AGACTCACAG TAGTTCTACG
951 CCAGTACAGC ACCCATCAAA ACCAGTGGTT CATCCAACGT CTACCCCAAG
1001 CACTGTTCTT TCTAGTCCAT TTACGCTACA GTCTGATCAC CAGCCAAAGA
1051 AATCATTTGA TGCTAATGGA GCATCTACTT TATCAAACT GCCTACACCC
1101 ACATCTTCTG TCCCTGCACA GAAACAGAAA AGAAAAGAA CTACATCAGG
1151 AGACAAACCC GTATCACATT CTTGCACAA TCCTTCCAGC TCTTCTGCCT
1201 CTGGAAGTAA CCCCACATCT GCACCTCCAA CATCTGCTTC AGCGGTCCCT
1251 GTTCTCTCTG TTCCACAGTC GCCAATACCT CCCTTACTTC AGGACCCAAA
1301 TCTTCTTAGA CAATTGCTTC CTGCTTTGCA AGCCACGCTG CAGCTTAATA
1351 ATTCTAATGT GGACATATCT AAAATAAATG AAGTTCTTAC AGCAGCTGTG
1401 ACACAAGCCT CACTGCAGTC TATAATTCAT AAGTTCTTAA CTGCTGGACC
1451 ATCTGCTTTC AACATAACGT CTCTGATTTT TCAAGCTGCT CAGCTCTCTA
1501 CACAAGATAT CCCTCTTCAT GAAGGTATCC AAATGGAGAG AGATACACAT
1551 AGGAGCAAAAT GGGAGTGAAG AGGGTCACCT TGTCAGAAAG CTGATAAACA
1601 GCAGGAATGC CTTGCTGGA ATGGAAGTAT AATGGTGCAA AGACTCTTGC
1651 AACCTCTGG CTAGCCTCAT GAGCAGGAGA CTGCGTGGA TACCTGGGCC
1701 TAAATGTAGA ATAAGAAAGA AGAAATAAGG ATGCCAGGCC ATCTAATCAG
1751 TCTCCAGTGT CTTTACATC TGATGCGTCA TCCCAAGAT CATATGTTTC
1801 TCCAAGAATA AGCACACCTC AAACATAACAC AGTCCCTATC AAACCTTTGA
1851 TCAGTACTCC TCCTGTTTCA TCACAGCCAA AGGTAGTATC TCCAGTAGTT
1901 AAGCAAGGAC CAGTGTACCA GTCAGCCACA CAGCAGCCTG TAACGTGCTGA
1951 CAAGCAGCAA GGTATGAAC CTGTCTCTCC TCGAAGCTTT CAGCGCTCAA
2001 GCCAGAGAAG TCCATCACCT GGTCCCAATC ATACTTCTAA TAGTAGTAAT
2051 GCATCAAAATG CAACAGTTGT ACCACAATAT TCTTCTGCCC GATCCACGTG
```

```
2101 TTCATTAACG CCTGCACTAG CAGCACACTT CAGTGAAAT CTCATAAAAC
2151 ACGTTCAAGG ATGGCCTGCA GATCATGCAG AGAAGCAGGC ATCAAGATTA
2201 CGGAAGAAG CGCATAACAT GGGAACTATT CACATGTCCG AAATTTGTAC
2251 TGAATTAATAA AATTTAAGAT CTTTAGTCCG AGTATGTGAA ATTCAAGCAA
2301 CTTTGGGAGA GCAAAGGATA CTATTTTGA GACAACAAAT TAAGGAACCT
2351 GAAAGCTAA AAAATCAGAA TTCCTTCATG GTGTGAAGAT GTGAATAATT
2401 GCACATGGTT TTGAGAACAG GAACTGTAAA TCTGTTGCC AATCTTAACA
2451 TTTTGGAGCT GCATTTAAGT AGACTTTGGA CCGTTAAGCT GGGCAAAGGA
2501 AATGACAAGG GGACGGGGTC TGTGAGATC AATTCAGGGG AAGATACAA
2551 GATTGATTG TAAACCCCTT GAAATGTAGA TTTCTGTAG ATGTATCCTT
2601 CACGTTGTAA ATATGTTTGT TAGAGTGAAG CCATGGGAAG CCATGTGTAA
2651 CAGAGCTTAG ACATCCAAAA CTAATCAATG CTGAGGTGGC TAAATACCTA
2701 GCCTTTTACA TGTAAACCTG TCTGCAAAAT TAGCTTTTTT AAAAAAAAAA
2751 AAAAAAAAAA AA
```

## BLAST Results

Entry AC005876 from database EMBLNEW:  
Homo sapiens chromosome 10 clone CIT987SK-118815 map 10p11.2-10p12.1,  
complete sequence.  
Score = 2130, P = 0.0e+00, identities = 426/426  
12 exons matching Bp 492-2740

## Medline entries

No Medline entry

## Peptide information for frame 2

ORF from 1757 bp to 2383 bp; peptide length: 209  
Category: questionable ORF  
Classification: no clue

```
1 MSLTSDASSP RSYVSPRIST POTNTVPIKP LISTPPVSSQ PKVSTPVVKQ
51 GPVSQSATQQ PVTADKQGGH EPVSPRSLQR SSQSPSPGPG NHTSNSSNAS
101 NATVVFPQNS ARSTCSLTPA LAHFSENLI KHVQGFADH AEKQASRLRE
151 EAHNMGTIHM SEICTELKNL RSLRVCEIQ ATLREQRILF LRQIKLEKEL
201 LKNQNSFMV
```

## BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphtes3\_19j17, frame 2

No Alert BLASTP hits found

## Peptide information for frame 3

ORF from 354 bp to 1661 bp; peptide length: 436  
Category: similarity to unknown protein  
Classification: unclassified  
Prosite motifs: WW DOMAIN\_1 (90-116)  
WW\_DOMAIN\_1 (90-116)

```
1 MRDAGDPSP NKMLRRSDSP ENKYSDSTCH SKAKNVHTR VRERDGGTSY
51 SPOENSHNHS ALHSSNSHSS NPSNPNPKTS DAPYDSADDW SEHISSSGKK
101 YYNCRTEVS QWEKPKEWLE REQROKEANK MAVNSFPKDR DYRREVMQAT
151 ATSGFASGME DKHSSDASSL LPQNILSQTs RHNDRDYRLP RAETHSSSTP
201 VQHPIKPVVH PTATPSTVPS SPFTLQSDHQ PKKSFDAANGA STLKSLPTPT
251 SSVPAQKTER KESTSGDKPV SHSCTTPSTS SASGLNPTSA PPTSASAVPV
301 SPVPOSPIPP LLQDPNLLRQ LLPALQATLQ LNNSNVDISK INEVLTAAVT
351 QASLQSIHK FLTAGPSAFN ITSLISQAAQ LSTQDIPLHE GIQMERDTHR
401 SKWEVKGSCL QKADKQEQCL VWNGSIMVQR LLQPSG
```

## BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphtes3\_19j17, frame 3

TREMBL:CEY40B1A.2 gene: "Y40B1A.2"; Caenorhabditis elegans cosmid  
Y40B1A, N = 1, Score = 144, P = 1.8e-09

>TREMBL:CEY40B1A.2 gene: "Y40B1A.2"; Caenorhabditis elegans cosmid Y40B1A  
Length = 120

HSPs:

Score = 144 (21.6 bits), Expect = 1.8e-09, P = 1.8e-09  
Identities = 30/67 (44%), Positives = 43/67 (64%)

Query: 90 WSEHSSSGKKYYNCRTEVSQWEKPKW-LEREQRQKEANKMAVNSFPK---DRDYRRE 145  
W+E +SSSGK YYN +TE+SQW+KP EW E +++ K VN P+ DR Y  
Sbjct: 11 WTEQMSSSGKMYNNKTEISQWDKPAEWPAEGGSAERDKPKGGVNEKPRFAEDR-YNEY 69

Query: 146 VMQATATS 153  
+ Q +++S  
Sbjct: 70 IGQLSSSS 77

Pedant information for DKFZphtes3\_19j17, frame 2

Report for DKFZphtes3\_19j17.2

[LENGTH] 209  
[MW] 22873.85  
[pI] 9.95  
[KW] All Alpha  
[KW] LOW\_COMPLEXITY 13.40 %

SEQ MSLTSDASSPRSYVSPRISTPQTNTVPIKPLISTPPVSSQPKVSTPVVKQGPVQSATQO  
SEG .....  
PRD ccc  
SEQ PVTADKQGGHEPVSPRSLQRSSQSPSPGPNHTSNSSNASNATVVPQNSSARSTCSLTPA  
SEG .....XX  
PRD cchh  
SEQ LAAHFSENLIKHVQGWPADHAKEQASRLREEAHNMGTIHMSEICTELKNLRLSLRVCEIQ  
SEG .....  
PRD hhhhhhchhhhhhhccccchhhhhhhhhhhhhhhccccchhhhhhhhhhhhhhhhhhhhh  
SEQ ATLREQRILFLRQQIKELEKLNQNSFMV  
SEG .....  
PRD hhhhhhchhhhhhhhhhhhhhhhhhhhhhhcccccccc

(No Prosite data available for DKFZphtes3\_19j17.2)

(No Pfam data available for DKFZphtes3\_19j17.2)

Pedant information for DKFZphtes3\_19j17, frame 3

Report for DKFZphtes3\_19j17.3

[LENGTH] 436  
[MW] 47716.62  
[pI] 8.71  
[HOMOL] TREMBL:CEY40B1A.2 gene: "Y40B1A.2"; Caenorhabditis elegans cosmid Y40B1A 6e-08  
[FUNCAT] 04.05.03 mrna processing (splicing) [S. cerevisiae, YKL012w] 2e-04  
[FUNCAT] 30.10 nuclear organization [S. cerevisiae, YKL012w] 2e-04  
[FUNCAT] 99 unclassified proteins [S. cerevisiae, YPR152c] 6e-04  
[BLOCKS] BL01159 WW/rsp5/WWP domain proteins  
[PROSITE] WW\_DOMAIN\_1 2  
[PFAM] WW/rsp5/WWP domain containing proteins  
[KW] All Alpha  
[KW] LOW\_COMPLEXITY 22.48 %



```
SEQ MRDAGDPSPFNKMLRRSDSPENKYS DSTGH SKAKNVHTRVREDDGGTSYSPQENSHNHS
SEG .....
PRD ccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc

SEQ ALHSSNSHSSNPSNNPSKTS DAPYDSADDWSEHISSSGKKYYNCRTEVSQWEKPKWELE
SEG xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx
PRD cccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccchhh

SEQ REQRQKEANKMAVNSFPKDRDYRREVMQATATSGFASGMEDKHSSDASSLLPQNILSQT
SEG .....
PRD hhhhhhhhhhhhhccccccccchhhhhhhhhhhcccccccccccccccccccccccccccccc

SEQ RHNDRDYRLPRAETHSSSTPVQHPIKPVVHPTATPSTVPSSPFTLQSDHQP KKSFDANGA
SEG .....
PRD ccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc

SEQ STLKSLPTPTSSVPAQKTERKESTSGDKPVSHSCTTPTSSASGLNPTSAPPTSASAVPV
SEG xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx
PRD ccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc

SEQ SPVPQSPIPPLLQDPNLLRQLPALQATLQLNNSNVDISKINEVLTA AVTQASLQSIHK
SEG xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx
PRD cccccccccccccccccchhhhhhhhhhhhhccccccccccccchhhhhhhhhhhhhhhhhhh

SEQ FLTAGPSAFNITSLISQAQLSTQDIPLHEGIQMERDTHRSKWEVKGSLCQKADKQEQECL
SEG .....
PRD hhccccccceehhhhhhhhhhhccccccccccccccccccccccccccccchhhhhhhccce

SEQ VWNGSIMVQRLQPSG
SEG .....
PRD eccccchhhhhcccccc
```

## Prosites for DKFZphtes3\_19j17.3

```
PS01159 90->116 WW_DOMAIN_1 PDOC50020
PS01159 90->116 WW_DOMAIN_1 PDOC50020
```

## Pfam for DKFZphtes3\_19j17.3

```
HMM_NAME WW/resp5/WWP domain containing proteins
HMM *LPsGWEeHWDpsGrpWYYWNHETKTTQWEpP*
+ ++W EH++ SG+ YY+N T+ +QWE+P
Query 86 SADDWSEHISSSGKK-YYNCRTEVSQWEKP 115
```

DKF2phtes3\_lcl

group: signal transduction

DKF2phtes3\_lcl encodes a novel 632 amino acid putative GTPase-activating protein, related to drosophila rotund transcript and human n-chimaerin.

rac small GTPase is associated with type-I phosphatidylinositol 4-phosphate 5-kinase and regulating the production of phosphatidylinositol 4,5-bisphosphate. The new protein is expected to activate p21rac-related small GTPases.

The new protein can find application in modulating/blocking the response to a cellular receptor.

similarity to GTPase-activating proteins

complete cDNA, complete cds, EST hits

Sequenced by DKF2

Locus: unknown

Insert length: 3237 bp

Poly A stretch at pos. 3227, no polyadenylation signal found

```

1  GCGAAGTGAA  GGGTGGCCCA  GGTGGGCGCA  GGCTGACTGA  ATGTATCTCC
51  TAGCTATGGA  CTAATAATA  CATGGGGGGA  AATAACAAG  TATTCATGAG
101  GGTGAAAATG  TGACCCAGCA  GGAAAATTAC  AACTATTTTC  AATTGACGTT
151  GAATAGGATG  AGTCATGGAA  TTTAAGTGAT  TTAATGAAGA  TTATACTACT
201  GGTAGATAGA  AGAGCTAAAG  AAAGATGGAT  ACTATGATGC  TGAATGTGCG
251  GAATCTGTTT  GAGCAGCTTG  TGCGCCGGGT  GGAGATTCTC  AGTGAAGGAA
301  ATGAAGTCCA  ATTTATCCAG  TTGGCGAAGG  ACTTTGAGGA  TTCCCGTAAA
351  AAGTGGCAGA  GGACTGACCA  TGAGCTGGGG  AAATACAAGG  ATCTTTTGAT
401  GAAAGCAGAG  ACTGAGCGAA  GTGCTCTGGA  TGTAAAGCTG  AAGCATGCAC
451  GTAATCAGGT  GGATGTAGAG  ATCAAACGGA  GACAGAGAGC  TGAGGCTGAC
501  TGCGAAAAGC  TGGAACGACA  GATTCAAGCTG  ATTCGAGAGA  TGCTCATGTG
551  TGACACATCT  GGCAGCATT  AACTAAGCGA  GGAGCAAAA  TCAGCTCTGG
601  CTTTCTCAA  CAGAGGCCAA  CCATCCAGCA  GCAATGCTGG  GAACAAGAAG
651  CTATCAACCA  TTGATGAATC  TGGTTCATT  TTAACAGATA  TCAGCTTTGA
701  CAAGACTGAT  GAATCACTGG  ATTGGGACTC  TTCTTGCTG  AAGACTTTCA
751  AACTGAAGAA  GAGAGAAAAG  AGCGCTCTTA  CTAGCCGACA  GTTTGTTGAT
801  GGTCCTCTG  GACCTGTAAA  GAAACTCGCT  TCCATTGGCT  CTGCAGTAGA
851  CCAGGGGAT  GAATCCATAG  TTGCAAAAAC  TACAGTGACT  GTTCCCAATG
901  ATGGCGGGCC  CATCGAAGCT  GTGTCCAATA  TTGAGACTGT  GCCATATTGG
951  ACCAGGAGCC  GAAGGAAAAC  AGGTACTTTA  CAACCTTGG  ACAGTGACTC
1001  CACCCTGAAC  AGCAGGCAGC  TGGAGCCAAG  AACTGAGACA  GACAGTGTGG
1051  GCACGCCACA  GAGTAATGGA  GGGATGCGCC  TGCTGACTTT  TGTTCCTAAG
1101  ACGGTTATTA  AACCTGAATC  CTGTGTTCCA  TGTGGAAGC  GGATAAAATT
1151  TGGCAATTA  TCTCTAAGT  GTGAGACTGT  TCGTGTGGTC  TCTCATCCAG
1201  AATGTCGGGA  CCGCTGTCCC  CTTCCCTGCA  TTCTACCCCT  GATAGGAACA
1251  CCTGTCAAGA  TTGAGAGGGG  AATGCTGGCA  GACTTTGTGT  CCCAGACTTC
1301  TCCAATGATC  CCCTCCATTG  TTGTGATTG  TGTAAATGAG  ATTGAGCAAA
1351  GAGGTCTGAC  TGAGACAGGC  CTGTATAGGA  TCTCTGGCTG  TGACCGCACA
1401  GTAAAAGAGC  TGAAGAGAG  ATTCTCAGA  GTGAAAACCT  TACCCTCTCT
1451  CAGCAAGAGC  GATGATATCC  ATGCTATCTG  TAGCCTTTCT  AAAGACTTTT
1501  TTCGAAACCT  CAAAGAACCT  CTTCTGACCT  TTCGCTTAA  CAGAGCCTTT
1551  ATGGAAGCAG  CAGAAATCAC  AGATGAAGAC  AACAGCATAG  CTGCCATGTA
1601  CCAAGCTGTT  GGTGAATGCG  CCCAGGCCAA  CAGGGACACA  TTAGCTTTCC
1651  TCATGATTCA  CTGACAGAGA  GTGGCTCAGA  GTCCACATAC  TAAAATGGAT
1701  GTTGCCAATC  TGGCTAAAGT  CTTTGGCCCT  ACAATAGTGG  CCCATGCTGT
1751  GCCCAATCCA  GACCCAGTGA  CAATGTTACA  GGACATCAAG  CGTCAACCCA
1801  AGGTGGTTGA  GCGCTGCTT  TCCTTGCCCT  TGGAGTATTG  GAGTCAAGTC
1851  ATGATGGTGG  AGCAAGAGAA  CATTGACCCC  TCATAGTCTA  TTGAAAACCT
1901  AAATGCCCTT  TCAACACCAC  AGACACCAGA  TATTAAGTGG  AGTTTACTGG
1951  GACCTGTGAC  CACTCTGAA  CATCAGCTTC  TCAAGACTCC  TTCTCTAGAT
2001  TCCCTGTAC  AGAGAGTCCG  TTCCACCCCT  ACCAAGAACA  CTCTAGATT
2051  TGGGAGCAAA  AGCAAGTCTG  CCACTAACCT  AGGACGACAA  GGCAACTTTT
2101  TTGCTTCTCC  AATGCTCAAG  TGAAGTCACA  TCTGCTGTT  ACTTCCCAGC
2151  ATTGACTGAC  TATAAGAAAG  GACACATCTG  TACTCTGCTC  TGCAGCCTCC
2201  TGTACTATT  ACTACTTTTA  GCATTCTCCA  GGCTTTTACT  CAAGTTTAAT
2251  TGTGATGAG  GGTTTTATTA  AACTATATA  TATCTCCCTC  TCCTTCTCCT
2301  CAAGTCACAT  AATATCAGCA  CTTTGTGCTG  GTCAATGTTG  GGAGCTTTTA
2351  GATGAGACAT  CTTTCCAGGG  GTAGAAGGGT  TAGTATGGAA  TTGGTTGTGA
2401  TTCTTTTGG  GGAAGGGGGT  TATTGTTCCT  TTGGCTTAAA  GCCAATTCCT
2451  GCTCATAGAA  TGATCTTTCT  CTAGTTTCAT  TAGAAGCTGA  TTCCGCTGAC
2501  ACAATGACAG  AAACCTTACC  TATCTGATA  GATTAGCTTG  TCTCAGGGTG
2551  GGAAGTGGGA  GGGCAGGGCA  AAGAAAGGAT  TAGACCAGAG  GATTTAGGAT

```

```

2601 GCCTCCTTCT AAGAACCAGA AGTTCTCATT CCCCATTTATG AACTGAGCTA
2651 TAATATGGAG CTTTCATAAA AATGGGATGC ATTGAGGACA GAACTAGTGA
2701 TGGGAGTATG CGTAGCTTTG ATTGGGATGA TTAGGCTTTT AATAGTGTG
2751 AGTGGGACAA CCTGTAAAT GTGAAAGTAC AACTGCTATT TATCTCTGAT
2801 GTGCGGCTGG CTGAACCTTG GTTTCATTGG GGCTCAAAGC CAGTTTTTCT
2851 TTTAAAAATTG AATTCATTCT GATGCTTGGC CCCCATACCC CCAACCTTGT
2901 CCACTGGAGC CCAACTTCTA AAGCTCAATA TATCATCCTT TGGCATCCCA
2951 ACTACAATA AAGAGTAGGC TATANGGAA GATTGTCAAT ATTTTGTGGT
3001 AGAAAAAGCT ACAGTCATTT TTTCTTTGCA CTTTGGATGC TGAATTTTTT
3051 CCCATGGAAC ATAGCCACAT CTAGATAGAT GTGAGCTTTT TCTTCTGTTA
3101 AAATTATTCT TAATGCTCTG TAAAACGATT TTCTTCTGTA GAATGTTTGA
3151 CTTCGTATTG ACCCTTATCT GTAAACACC TATTTGGGAT AATATTGGA
3201 AAAAAAGTAA ATAGCTTTTT CAAAATGAAA AAAAAA

```

## BLAST Results

Entry U82984 from database EMBLEST:  
Homo sapiens DRES 56 mRNA sequence.  
Score = 8775, P = 0.0e+00, identities = 1757/1758  
matches 3' end

## Medline entries

93074974:  
Developmental regulation and neuronal expression of the mRNA of rat  
n-chimaerin, a  
p21rac GAP:cDNA sequence.

93024458:  
A Drosophila rotund transcript expressed during spermatogenesis and  
imaginal disc  
morphogenesis encodes a protein which is similar to human Rac  
GTPase-activating  
(racGAP) proteins.

## Peptide information for frame 3

ORF from 225 bp to 2120 bp; peptide length: 632  
Category: similarity to known protein

```

1 MDTMMLNVRN LFEQLVRRVE ILSEGNEVQF IQLAKDFEDF RKKWQRTDHE
51 LGKYKDLLMK AETERSALDV KLKHARNQVD VEIKRRQRAE ADCEKLERQI
101 QLIREMLMCD TSGSIQLSEE QKSALAFNLR GQPSSSNAGN KRLSTIDESG
151 SILSDISFDK TDESLOWDSS LVTFKLLKKR EKKRSTSRQF VDGPPGPVKK
201 TRSIGSAVDQ GNESIVAKTT VTPVNDGGPI EAVSTIETVP YWTRSRRTKG
251 TLQFWNSDST LNSRQLEPRT ETDSVGTQPS NGGMRLHDFV SKTVIKPESC
301 VPCGKRIKFG KLSLKRCDRC VVSHPECRDR CPLPCIPTLI GTPVKIGEGM
351 LADFVSQTSF MIPSIVVHCV NEIEQRGLTE TGLYRISGCD RTVKELKEKF
401 LRVKTVPLLS KVDDIHAICS LLKDFLRNLK EPLLTFRNLN AFMEAAEITD
451 EDNSIAAMYQ AVGELPQANR DTLAFLMIHL QRVAQSPHTK MDVANLAKVF
501 GPTIVAHAVP NPDPTMLQD IKRQPKVVER LLSLPLEYWS QFMVVEQENI
551 DPLHVIENSN AFSTPQTPOI KVSLLGPVTT PEHQLLKTPS SSSLSQRVRS
601 TLTKNTPRFG SKSKSATNLG RQGNFFASPM LK

```

## BLASTP hits

Entry CEK08E3.4 from database TREMBLNEW:  
gene: "K08E3.6"; Caenorhabditis elegans cosmid K08E3  
Score = 452, P = 2.6e-48, identities = 126/377, positives = 189/377

Entry A48122 from database PIR:  
GTPase-activating protein Rac homolog, splice form clone pcl.7 - fruit  
fly (Drosophila melanogaster) (fragment)  
Score = 480, P = 9.2e-46, identities = 111/270, positives = 155/270

Entry B48122 from database PIR:  
GTPase-activating protein Rac homolog, splice form clone pcl.7d - fruit  
fly (Drosophila melanogaster)  
Score = 480, P = 9.2e-46, identities = 111/270, positives = 155/270

Entry DM22539\_1 from database TREMBL:  
gene: "rotund"; product: "rnracGAP"; Drosophila melanogaster rnracGAP  
(rotund) gene, complete cds.  
Score = 480, P = 9.2e-46, identities = 111/270, positives = 155/270

Entry S29128 from database PIR:  
N-chimerin - rat  
Score = 336, P = 8.8e-30, identities = 86/253, positives = 128/253

Alert BLASTP hits for DKF2phtes3\_lcl, frame 3

No Alert BLASTP hits found

Pedant information for DKF2phtes3\_lcl, frame 3

Report for DKF2phtes3\_lcl.3

[LENGTH] 632  
[MW] 71026.84  
[PI] 9.08  
[HOMOL] PIR:B48122 GTPase-activating protein Rac homolog, splice form clone pcl.7d -  
fruit fly (Drosophila melanogaster) 2e-46  
[FUNCAT] 10.99 other signal-transduction activities [S. cerevisiae, YBR260c] 3e-12  
[FUNCAT] 03.22 cell cycle control and mitosis [S. cerevisiae, YER155c] 2e-11  
[FUNCAT] 30.03 organization of cytoplasm [S. cerevisiae, YER155c] 2e-11  
[FUNCAT] 03.04 budding, cell polarity and filament formation [S. cerevisiae, YER155c]  
2e-11  
[FUNCAT] 03.10 sporulation and germination [S. cerevisiae, YDL240w] 3e-09  
[FUNCAT] 30.04 organization of cytoskeleton [S. cerevisiae, YOR134w] 4e-09  
[FUNCAT] 06.10 assembly of protein complexes [S. cerevisiae, YOR134w] 4e-09  
[FUNCAT] 03.07 pheromone response, mating-type determination, sex-specific proteins  
[S. cerevisiae, YOR127w] 5e-09  
[FUNCAT] 09.04 biogenesis of cytoskeleton [S. cerevisiae, YPL115c] 3e-08  
[FUNCAT] 10.02.09 regulation of g-protein activity [S. cerevisiae, YPL115c] 3e-08  
[BLOCKS] BL00479B Phorbol esters / diacylglycerol binding domain proteins  
[BLOCKS] BL00479A Phorbol esters / diacylglycerol binding domain proteins  
[SCOP] dlpbwa\_1.83.1.1.2 p85 alpha subunit RhoGAP domain [human (Homo sapiens)] 1e-55  
[SCOP] dirgp\_1.83.1.1.1 p50 RhoGAP domain [human (Homo sapiens)] 1e-49  
[PIRKB] breakpoint cluster region 1e-19  
[PIRKB] transmembrane protein 7e-08  
[PIRKB] brain 3e-22  
[PIRKB] alternative splicing 1e-19  
[PIRKB] P-loop 2e-25  
[SUPFAM] CDC24 homology 3e-22  
[SUPFAM] bcr protein 3e-22  
[SUPFAM] myosin motor domain homology 2e-25  
[SUPFAM] pleckstrin repeat homology 4e-10  
[SUPFAM] LIM metal-binding repeat homology 2e-09  
[SUPFAM] protein kinase C zinc-binding repeat homology 5e-29  
[PROSITE] MYRISTYL 6  
[PROSITE] AMIDATION 1  
[PROSITE] CAMP\_PHOSPHO\_SITE 3  
[PROSITE] CK2\_PHOSPHO\_SITE 13  
[PROSITE] TYR\_PHOSPHO\_SITE 2  
[PROSITE] PKC\_PHOSPHO\_SITE 9  
[PROSITE] ASN\_GLYCOSYLATION 1  
[PROSITE] DAG\_PE\_BINDING\_DOMAIN 1  
[PFAM] Phorbol esters / diacylglycerol binding domain  
[KW] Irregular  
[KW] 3D  
[KW] LOW\_COMPLEXITY 2.22 %  
[KW] COILED\_COIL 8.54 %

SEQ MDTMMLNVRNLFQQLVRRVEILSEGNEVQFIQLAKDFEDFRKKWQRTDHELKGYKDLLMK  
SEG .....  
COILS .....CCCCCCCCCCCC  
lrqp-.....  
SEQ AETERSALDVKLKHARNQVDVEIKRRQRAEADCEKLERQIQLIREMLMCDTSGSILQSEE  
SEG .....  
COILS CC  
lrqp-.....  
SEQ QKSALAFNLNRGPSSSNAGNKLSTIDSGSILSDISFDKTDSELDWSSLVKTFKLLKKR  
SEG .....  
COILS .....

```

lrqp- .....
SEQ   EKRRSTSRQFVDGPPGPVKKTRSIGSAVDQGNESIVAKTTVTVPNDGGPIEAVSTIETVP
SEG   .....
COILS .....
lrqp- .....

SEQ   YWTRSRRTGTTLQPWNSDSTLNSRQLEPRTETDSVGT PQSNGGMRLHDFVSKTVIKPESC
SEG   .....
COILS .....
lrqp- .....

SEQ   VPCGKRIKFGKLSLKCRDCRVVSHPECRDRCLPCIPTLIGTPVKIGEGMLADFVSQTSP
SEG   .....
COILS .....
lrqp- .....

SEQ   MIPSIVVHCVNEIEQRGLTETGLYRISGCDRTVKELKEKFLRVKTVPLLSKVDDIHAICS
SEG   .....
COILS .....
lrqp- .CCHHHHHHHHHHHHHHHHTTTTTTTTCCCHHHHHHHHHHHHHHCCCCG-GGCCCCHHHHH

SEQ   LLKDFLRNLKEPLLTFRNRAFMEAAEITDEDNSIAAMYQAVGELPQANRDTLAFMIHL
SEG   .....
COILS .....
lrqp- HHHHHHHHTTTTTTGGGHHHHHHHTTT-CGGGHHHHHHHHHHHCCCHHHHHHHHHHHHHH

SEQ   QRVAQSPTKMDVANLAKVFGPTIVAHAVPNPDVMTLQDIKRQPKVVERLLSLPLEYWS
SEG   .....
COILS .....
lrqp- HHHHHHHHCCCHHHHHHHHGGGCC.....

SEQ   QFMVVEQENIDPLHVIENSNAFSTPQTPDIKVSLLGPVTTPEHQLKTPSSSSLSQRVRS
SEG   .....
COILS .....
lrqp- .....

SEQ   TLTKNTPRFGSKSKSATNLGRQGNFFASPMLK
SEG   xxx.....
COILS .....
lrqp- .....

```

## Prosites for DKFZphtes3\_lcl.3

PS00001	212->216	ASN GLYCOSYLATION	PDOC00001
PS00004	141->145	CAMP_PHOSPHO_SITE	PDOC00004
PS00004	182->186	CAMP_PHOSPHO_SITE	PDOC00004
PS00004	246->250	CAMP_PHOSPHO_SITE	PDOC00004
PS00005	63->66	PKC_PHOSPHO_SITE	PDOC00005
PS00005	174->177	PKC_PHOSPHO_SITE	PDOC00005
PS00005	186->189	PKC_PHOSPHO_SITE	PDOC00005
PS00005	245->248	PKC_PHOSPHO_SITE	PDOC00005
PS00005	313->316	PKC_PHOSPHO_SITE	PDOC00005
PS00005	392->395	PKC_PHOSPHO_SITE	PDOC00005
PS00005	435->438	PKC_PHOSPHO_SITE	PDOC00005
PS00005	595->598	PKC_PHOSPHO_SITE	PDOC00005
PS00005	606->609	PKC_PHOSPHO_SITE	PDOC00005
PS00006	47->51	CK2_PHOSPHO_SITE	PDOC00006
PS00006	66->70	CK2_PHOSPHO_SITE	PDOC00006
PS00006	144->148	CK2_PHOSPHO_SITE	PDOC00006
PS00006	206->210	CK2_PHOSPHO_SITE	PDOC00006
PS00006	234->238	CK2_PHOSPHO_SITE	PDOC00006
PS00006	270->274	CK2_PHOSPHO_SITE	PDOC00006
PS00006	323->327	CK2_PHOSPHO_SITE	PDOC00006
PS00006	387->391	CK2_PHOSPHO_SITE	PDOC00006
PS00006	392->396	CK2_PHOSPHO_SITE	PDOC00006
PS00006	410->414	CK2_PHOSPHO_SITE	PDOC00006
PS00006	449->453	CK2_PHOSPHO_SITE	PDOC00006
PS00006	489->493	CK2_PHOSPHO_SITE	PDOC00006
PS00006	579->583	CK2_PHOSPHO_SITE	PDOC00006
PS00007	46->55	TYR_PHOSPHO_SITE	PDOC00007
PS00007	376->385	TYR_PHOSPHO_SITE	PDOC00007
PS00008	131->137	MYRISTYL	PDOC00008
PS00008	150->156	MYRISTYL	PDOC00008
PS00008	276->282	MYRISTYL	PDOC00008
PS00008	377->383	MYRISTYL	PDOC00008
PS00008	388->394	MYRISTYL	PDOC00008
PS00008	623->629	MYRISTYL	PDOC00008
PS00009	303->307	AMIDATION	PDOC00009

WO 01/12659

PCT/IB00/01496

PS00479 287->336 DAG\_PE\_BINDING\_DOMAIN PDOC00379

Pfam for DKF2phtes3\_lcl.3

HMM_NAME	Phorbol esters / diacylglycerol binding domain		
HMM	*HrFmrHTFrqPTWCDHCgeFIWGwgKQGYQCQnCGmNCHKRCHe1VPmm		
	H+E+ +T + P +C CG +I +GK ++C +C+++ H +C+ + P		
Query	287	HDFVSKTVIKPESCVPCKRI-KFGKLSLKRCRVRVSHPECRDRCPLP	334
HMM	C*		
	C		
Query	335	C	335

DKF2phtes3\_lg13

group: intracellular transport and trafficking

DKF2p DKF2phtes3\_lg13 encodes a novel 1007 amino acid protein with similarity to human 256 kD golgin.

The new protein contains 7 leucine zippers and seems to be involved in protein-protein-interaction in the golgi apparatus. The very similar rat cpl51 shows haploid-specific transcription in mus musculus testis.

The new protein can find application in modulating protein traffic in the golgi apparatus, especially in human haploid germ cells.

similarity to 256 kD golgi, strong similarity to rat "cpl51"

21 exons encoded on AC004682

EST from a testis library, two mouse ESTs of a testis cDNA library, rat cpl51 shows haploid-specific transcription!  
testis or haploid-specific transcription

Sequenced by DKF2

Locus: map="16q22.2"

Insert length: 3405 bp

Poly A stretch at pos. 3394, polyadenylation signal at pos. 3373

```

1 GGGATAGGGG ATGTGGTTG TTACAAAGGA TGAGTATTTT GATAGCTTCT
51 CATTCTCTGA ACTATTCTGC AGGTTTATAA CAAAGCTCAG AAAATACTAA
101 AGGTTAAAGG AGAATTGAGA GCTGCCAAGG AAATGAAAGA TGAGCGGGGG
151 GAGAGAGACA GAGAAGTGAG CAGCCTGAAC AGCAAGCTGT TAAGCCTGCA
201 ACTTGACATC AAGAATCTGC ACGATGTCTG CAAGAGACAG AGGAAGACCT
251 TGCAGGACAA TCAGCTCTGC ATGGAGGAGG CAATGAACAG CAGCCACGAC
301 AAGAAGCAAG CACAGGCATT AGCATTGAGG GAGTCAGAGG TGGAAATTTGG
351 GTCCAGTAAA CAGTGTCTATC TGAGACAACCT CCAGCAACTG AAGAAAAAAT
401 TGCTGGTGCT TCAACAAGAA CTGGAGTTTC ACACAGAGGA GTTGACAGCT
451 TCTTACTATT CTCTCCGCCA GTATCAGTCC ATCTAGAGA AGCAGACTTC
501 CGACCTGGTT CTCTCGACCC ATCACTGCAA ACTGAAAGAA GATGAGGTGA
551 TTCTCTATGA GGAGGAAATG GGAAATCACA ACGAGAACAC AGGGGAGAGC
601 CTCCATTTGG CGCAGGAGCA ACTCGCCTTG GCCGGGGACA AGATCGCCTC
651 TCTAGAGAGG AGCTTAAAGC TCTACAGGGA TAAATACCCAG TCTTCCCTGA
701 GCAACATCGA GTTACTAGAA TGCCAAGTGA AGATGTTTGA GGGGGAACCT
751 GGGGGATACA TGGGTACAGG GCCTCAGAAC AAGGCTGATC ATTCAAAGGT
801 ACGGATATAC ACTTCTCCTT GCATGATTCA AGAGCATCAG GAGACTCAGA
851 AACGACTGTC TGAAGTCTGG CAAAAGGTCT CTCACAGGA TGATCTCAIT
901 CAAGAAGTTC GAAATAAGCT GGCCTGCAGT AACGCTTTGG TTCTGGAGCG
951 TGAAAAGGCT TTGATAAATC TACAAGCCGA TTTTGCTTCC TGTACAGCCA
1001 CCCACAGATA CCCTCTAGC TCCTCAGAAG AGTGTGAAGA CATCAAAAAG
1051 ATACTGAAGC ACTTGACGGA GCAGAAAGAC AGCCAGTGCC TGCATGTGGA
1101 GGAGTACCAG AACCTGGTGA AGGATCTGCG CGTGGAACTA GAGGCGGTGT
1151 CGGAACAGAA GAGAAACATC ATGAAGGACA TGATGAAGCT GGAGCTGGAC
1201 CTGCACGGAC TCGGGGAGGA GACATCTGCC CACATTGAGA GGAAGGATAA
1251 GGACATCACC ATCTGCGAGT GCCGGCTGCA GGAGCTGCAG CTGAGATTCA
1301 CCGAGACCCA AAAGCTCACT TTGAAGAAAG ACAAGTTCTT CCAAGAGAAA
1351 GATGAGATGC TGCAAGAGCT GGAGAAGAAA CTGACACAGG TTCAGAACAG
1401 CCTCCTGAAA AAGGAGAAGG AGCTGGAGAA GCAGCAGTGC ATGGCCACAG
1451 AACTTGAAAT GACAGTCAAG GAGGCTAAGC AGGACAAATC CAAGGAGGCG
1501 GAGTGCAAGG CCCTGCAGGC TGAGGTCCAG AAGCTGAAGA ACAGTCTCGA
1551 AGAGGCCAAG CAGCAGGAGA GGCTGGCTGC TCAGCAAGCA GCCCAGTGCA
1601 AAGAAGAGGC TGCACTGGCA GGCTGTCCAC TGGAGGACAC CCAGAGGAAA
1651 CTGCAGAAGG GTCTCTCCTT GGACAAGCAG AAGGCAGACA CCATCCAGGA
1701 ACTACAGAGA GAACCTCAGA TGCTGCAGAA GGAGTCTCGT ATGGCTGAGA
1751 AGGAACAACAC CTCCAACAGA AAACGGGTGG AGGAGCTGTC ATTAGAACTC
1801 TCTGAAGCCC TGAGGAAGCT TGAATAATTA GACAAGGAAA AGAGGCAGCT
1851 TCAGAAGACA GTGGCTGAGC AGGATATGAA AATGAATGAC ATGCTTGATC
1901 GTATCAAGCA CCAGCAGAGG GAGCAAGGCT CCATCAATG CAAGTTAGAA
1951 GAAGATCTTC AGGAGGCCAC AAAGCTTCTG GAGGACAAAC GGGAGCAGTT
2001 GAAGAAGAGC AAAGAGCATG AGAAGCTGAT GGAGGGAGAA CTTGAAGCTT
2051 TCGGGCAGGA ATTTAAAAAG AAAGACAAAG CGTTGAAAGA GAATTCAGAA
2101 AAGTTGAGGG AAGAAAATGA GAATCTCCGA GCAGAGCTAC AGTGTGTTTC
2151 TACACAACTG GAATCTCTC TCAACAAATA CAACACCAGC CAGCAAGTCA
2201 TCCAAGACTT GAATAAAGAG ATAGCCCTTC AGAAGGAGTC CTTAATGAGC
2251 CTGACAGGCC AGCTGGACAA AGCTCTGAGC AAGGAGAAAG ACTATCTCCA
2301 GACTACCATC ACCAAAGAAAG CCTATGATGC ATATCTCCGG AAGTCAGCCG
2351 CCTGCCAGGA TGACCTGACA CAAGCCCTCG AGAAGCTCAA TCACGTGACC
2401 TCAGAGACAA AGAGCTGCA GCAAAGCTTG ACACAGACCC AAGAGAAGAA

```

```
2451 AGCTCAGCTG GAAGAGGAAA TCATTGCTTA TGAGGAAAGG ATGAAAAAGC
2501 TCAATACGGA ATTAAGAAAA CTGCGGGGCT TCCACCAGGA GAGTGAGCTG
2551 GAGGTGCACG CCTTTGACAA GAAGCTAGAG GAGATGAGCT GCCAGGTGCT
2601 GCAGTGGCAG AAGCAACACC AGAATGACCT CAAGATGCTG GCAGCCAAAG
2651 AGGAGCAGCT CAGGGAGTTC CAGGAGGAGA TGGCCGCCTT AAAAGAGAAC
2701 CTCCTTGAGG ACGATAAGGA GCCCTGCTGC CTGCCCCAGT GGTCTGTGCC
2751 CAAAGACACC TGTAGGCTCT ACCGAGGGAA TGATCAGATT ATGACCAACT
2801 TGGAGCAATG GGCAAAACAG CAGAAGGTGC CCAATGAGAA ACTAGGAAAC
2851 CAGCTCCGAG AGCAGGTGAA CTACATTGCC AAGCTGAGTG GCGAAAAGGA
2901 CCACCTCCAC AGTGAATGG TCCACTTGCA GCAGGAAAC AAGAGCTGA
2951 AGAAGGAGAT AGAAGAGAAG AAGATGAAG CCGAGAACAC AAGGCTATGC
3001 ACCAAGGCC TAGGCCCGAG CAGAACGGAG TCCACACAGA GGGAGAAAGT
3051 GTGGGGCACC TTGGGCTGGA AGGGGTGCCC CCAGGATATG GGTCAAAAGAA
3101 TGGACCTCAC CAAGTACATC GGGATGCCCC ACTGCCCGGG TTCCTCATAC
3151 TGCTAGAATC CACATCTAGC CCTGAGCAGC ATTTCCACGG GTGTTCTTTC
3201 AGAGGACAGT GAGTTCCAG CCCTCCCTCT CTCTTGACCT GGATCAGCTC
3251 TTACAGGAGT ATATCAGGGT CCCAGCCTAT TTTCGAAGAC ACTAATTTT
3301 GTTGAGTTTT GTCCACTTCC TGCCATGGAG TGAGCTTTAG AACCATACTA
3351 CCATCTCCAG GCCCAAACCT TGAATAAAG ACATGAGCAT GAGCAAAAAA
3401 AAAAA
```

## BLAST Results

Entry AC004682 from database EMBLNEW:  
Homo sapiens Chromosome 16 BAC clone CIT987SK-A-259H10, complete  
sequence.  
Score = 1291, P = 0.0e+00, identities = 265/272

## Medline entries

No Medline entry

## Peptide information for frame 1

ORF from 133 bp to 3153 bp; peptide length: 1007  
Category: similarity to known protein  
Prosite motifs: LEUCINE\_ZIPPER (83-105)  
LEUCINE\_ZIPPER (90-112)  
LEUCINE\_ZIPPER (97-119)  
LEUCINE\_ZIPPER (104-126)  
LEUCINE\_ZIPPER (403-425)  
LEUCINE\_ZIPPER (410-432)  
LEUCINE\_ZIPPER (918-940)

```
1 MKDEAGERDR EVSSLNSKLL SLQLDIKNLH DVCKRQRKTL QDNQLCMEEA
51 MNSSHDKKQA QALAFEESEV EFGSSKQCHL RQLQQLKKKL LVLQOELEFH
101 TEELQTSYYS LRQYSILEK QTSDLVLLHH HCKLKEDEVI LYEEMGNHN
151 ENTGELHLA QEQALAGDK IASLERSLNL YRDKYQSSLS NIELLECQVK
201 MLQELGGIM GQEPENKGDH SKVRIYTPC MIQEHQETQK RLSEVMQKVS
251 QQDDLIQELR NKLACSNALV LEREKALIKL QADFASTCT HRYPPSSSEE
301 CEDIKKILKH LQEQKDSQCL HVEEYQNLVK DLRVELEAVS EQKRNIMKDM
351 MKLELDLHGL REETSAHIER KDKDITILQC RLQELQLEFT ETQKLTLLKD
401 KFLQEKDEML QELEKKLTQV QNSLLKKEKE LEKQOCHATE LEHTVKEAQ
451 DKSKEAECKA LQAEVQKLN SLEAKQQR LAAQQAQCK EEAALAGCHL
501 EDTQKLEQK LLLDKQKADT IQELQRELOM LQKESMAEK EQTSNRKRVE
551 ELSLESEAL RKLENSDKEK RQLQKTVAEQ DMKMNMDLDR IKHQHREQGS
601 IKCKLEEDLQ EATKLEDKRL EQLKKSKEHE KLMEGELEAL QDLFKKKDKT
651 LKENSRLKEE ENENLRAELO CCSTOLESSL NKYNTSQQVI QDLNKEIALQ
701 KESLMSLQAO LDKALQKEKH YLQTTITKEA YDALSRKSA CQDDLTOALE
751 KLNHVTSYTK SLQSLTQTO EKKAQLEEEI IAYEERMKKL NTELRLKRGF
801 HQESELEVHA FDKKLEEMSC QVLQWQKQHQ NDLKMLAAKE EQLREFQEEM
851 AALKENLLED DKEPCCLPQW SVPKDTCLRY RGNDQIMTNL EQWAKQKVA
901 NEKLGNOLRE QVNYIAKLSG EKDLHLSVMV HLQOENKKLK KEIEEKKMKA
951 ENTRLCTKAL GPSRTESTQR EKVCGLGWK GLPQDMGQRM DLTKYIGMPH
1001 CPGSSYC
```

## BLASTP hits

Entry HS417401\_1 from database TREMBL:  
product: "trans-Golgi p230"; Human trans-Golgi p230 mRNA, complete



cds.  
Score = 411, P = 3.9e-34, identities = 212/862, positives = 420/862  
Entry SCINTANA\_1 from database TREMBL:  
Saccharomyces Cerevisiae integrin analogue gene, complete cds.  
Score = 404, P = 6.2e-34, identities = 199/897, positives = 423/897  
Entry HS6802\_2 from database TREMBL:  
gene: "MYH9"; product: "dJ6802.2"; Homo sapiens DNA sequence from PAC  
6802 on chromosome 22. Contains apolipoprotein L, myosin heavy chain,  
ESTs, CA repeat, STS and GSS.  
Score = 404, P = 1.9e-33, identities = 231/1028, positives = 469/1028  
Entry AF092090\_1 from database TREMBL:  
product: "cp151"; Rattus norvegicus cp151 mRNA, partial cds.  
Score = 2523, P = 3.0e-262, identities = 506/733, positives = 611/733

## Alert BLASTP hits for DKFZphtes3\_lgl3, frame 1

TREMBL:HS6802\_2 product: "256 kD golgin"; H.sapiens mRNA for golgin,  
N = 1, Score = 411, P = 4.4e-34

TREMBL:HS417401\_1 product: "trans-Golgi p230"; Human trans-Golgi p230  
mRNA, complete Cds., N = 1, Score = 411, P = 4.5e-34

TREMBL:SCINTANA\_1 Saccharomyces cerevisiae integrin analogue gene,  
complete cds., N = 1, Score = 404, P = 7.1e-34

>TREMBL:HS6802\_2 product: "256 kD golgin"; H.sapiens mRNA for golgin  
Length = 2,185

## HSPs:

Score = 411 (61.7 bits), Expect = 4.4e-34, P = 4.4e-34  
Identities = 212/816 (25%), Positives = 420/816 (51%)

Query: 145 EMGNHNEN-TGEKHLAQEQALAGDKIASLERSLNLYRDKYQSSLSNIELLECQVKMLQ 203  
+H + E+ G L +EQL ++ +ERSL+ YR KY ++ +L+ + K LQ  
Sbjct: 119 DMDSEADLVGNSDSLANKQLI---QRLRMERSLSSYRGKYSLVATYQMLQREKKKLQ 175  
Query: 204 GELGGIMGOEPENKGDHSHVRIYTPCMIQEHQETQKRLSEVWQ-KVSQDDLIQELRNK 262  
G I+ Q D S RI +Q Q+ +K L E + + +D I L+ +  
Sbjct: 176 G----ILSQSQ----DKSLRRIAREELQMDQQAQKHLQEEFDASLEEKQYISVLQTO 227  
Query: 223 LAC-----SNALVLEREKALIKLQADFASCTATHRYPPSSSEEC-ED--IKKILKLQE 313  
++ + + ++ K L +L+ A P S E ED K L+ LQ+  
Sbjct: 268 VSLLKQRLRNGPMNVDPVLPQLEPQ-AEVFTKEENPESDGPVVEDGTSVKTLETLOQ 286  
Query: 314 QKDSQ-----CLH-VEEYQNLVKDLRVELEAVSEQKRNIMKDMMLKLELDLHGLREETS 366  
+ Q C ++ ++ L E EA+ EQ +++ K+ DLH + E+T  
Sbjct: 287 RVKRQENLLKRCKETIQSHKEQCTLLTSEKEALQELDERLQLEKIK-DLH-MAEKTCL 344  
Query: 367 HIERKDKDITILQCLQELQLEFTEQKLTLLKDKFLQEKDEMLQLEKKLTQV--QNSL 424  
+ +D I Q Q+ + ET++ + + L+ K+E + +L ++ Q+ Q  
Sbjct: 345 ITQLRADAKNLIEQLE-QDKGMVIAETKR---QMHETLEMKEEIAQLRSRIKQMTTGEE 400  
Query: 425 LKKEKELEKQQCMATELEMTVKEAKQDKSKEAECKALQAEVQKLNKSLLEAKQGERLAAQ 484  
L+++KE + ++ ELE + A+ K++EA K L+AE+ + ++E+ ++ER++ Q  
Sbjct: 401 LREQKE-KSERAAFEELKALSTAQ--KTEEARRK-LKAEMDEQIKTIKTSEERISLQ 456  
Query: 485 QA-AQCNEEAA-LAGCHLEDTQKRLQGLLLQKQKADTIQELQELQMLQKSSMAEKEQ 542  
Q ++ K+E + E+ RLQR L +K+ A QEL ++LQ ++E E+ +  
Sbjct: 457 QELSRVQEVVDVMKSSSEEQIARLQK--LHEKELARKEQELTKKLQTRERE--FQEQMK 512  
Query: 543 TSNRRRVEELSLELSEALRKLKLENSDKEKRLQKT--VAEQDMKMNMDLRIKHQHQREQGS 600  
+ K E L++S+ +E+ E+ +LQK + E + K+ D+ +  
Sbjct: 513 VALEKSQSEY-LKISQKEQESLALLEELELQKKAILTESENKLRDLQQAETVTRILE 571  
Query: 601 IKCKLEEDLQEA TKLEED----KREQLKKSKEHEKLMEG---ELEALR-QEFKKKOKTL 651  
++ LE+ LQE +D + E+ K +KE ++E ELE+L+ Q+ + L  
Sbjct: 572 LESSLEKSLQENKQSKDLAVHLEAKNKHKEITVMVEKHKTELESKHKQDQALWTEKL 631  
Query: 652 KENSRKLEENENLRAELQCCSTQLESSL-NKYNTSQOVIQDLNKE----IALQKESLMS 706  
+ ++ + E E LR + C + E+ L +K Q I++N++ + ++ L S  
Sbjct: 632 QVLKQQYQTEMEKLREX---CEQEKETLLKDKKEIFQAHIEEMNEKTEKLDVQKQTELES 688  
Query: 707 LQAQLDKALQKEKHYLQT--YITKEAYDALSRKSAACQDDLTQALEKLNHVTSKSLQ 764  
L ++L + L K +H L+ ++ K+ D + ++ A D+ Q V S K +

Sbjct: 689 LSSELSEVL-KARHKLEELSVLKQDQTKMKQLEAKMDE--QKNHHQQQVDSIIKEHEV 745

Query: 765 SLTQTQEKKAQLEEEIIAYEERMKKLNTLRLGRGFHGESELEVHAFDCKLEEMSCQVLQ 824  
S+ +T+ KA L+++I E +K+ + L++ + + E ++ + +L++ S ++

Sbjct: 746 SIQRTE--KA-LKDQINQLELLKDKHKEHQAHVENLEADIKRSEGELOQASAKLDV 802

Query: 825 WQKQHNDLKMMLAAKEEQLEFQEEAALKENLLEDDKEPCCLPQW----SVPKDT-C-R 878  
+Q +Q+ A EQ + ++E++A L++ LL+ + E L + + KD C

Sbjct: 803 FQS-YQS-----ATHEQTKAYEEQLAQLQKLLDLETERILLTKQVAEVEAKKDVCTE 855

Query: 879 LYRGNDQIMTNLEQWAKQKQVANEKLGNLREQVNYIAKLS-GEKDLHLSVMVHLQOENK 937  
L Q+ ++Q KQ +K+ + QV Y +KL G K+ + + ++EN

Sbjct: 856 LDAHKIQVDLMQLEKQNSEMEQKVKSLT--QV-YESKLEDGNKEQETKQILVEKENM 912

Query: 938 KLK-KEIEEKKMAENTRLCTK 958  
L+ +E ++K+++ +L K

Sbjct: 913 ILQMRGQKKEIILTQKLSAK 934

Score = 338 (50.7 bits), Expect = 3.1e-26, P = 3.1e-26  
Identities = 216/953 (22%), Positives = 468/953 (49%)

Query: 2 KDEAGERDRE--VSSLNS-KLL-SLQLOIKNLHDVCKRQRKTLQDN-OLCM-----EEAM 51  
K+E E D E V S K L +LQ +K ++ KR ++T+O + + C +EA+

Sbjct: 260 KEENPESDCEPVVEDGTSVKTLETLOQRVKKRQENLLKRCKETIQSHKEQCTLLTSEKEAL 319

Query: 52 NSSHDKQAQALAFEESEVEFGSSKQCHLRQ----LQQLK--KKLLVLQLEFHETELQ 105  
D++ + ++ + + LR ++QL+ K +++ + + H E L+

Sbjct: 320 QEQLDERLQLEKIKDLHMAEKTKLITQLDAKNLIEQLEQDKGMVIAETKRQMH-ETLE 378

Query: 106 TSYSLRQYQSILEKQTSIDLVLHHHCKLKEDEVILYEEEMGNHNENTGEKHLAQEQL- 164  
+ Q +S +++ T+ L K K + E E +T+K A+ +L

Sbjct: 379 MKEEEIAQLRSRIKQMTTQGEELREQ-KEKSERAFAFELEKAL--STAQTTEARRKLK 434

Query: 165 ALAGDKIASLERSLNLRYDKYQSSLSNI--ELLEQVKMLQELGGIMQGEFENKGDHSK 222  
A ++I ++E++ R Q LS + E+++ K + ++ + Q+ K K

Sbjct: 435 REMDEQIKTIKTSEEEERISLQELSRVQEVVDVNMKSSEEQIAKL--QKLHEKELARK 492

Query: 223 VRIYTPCMIQEHQETQKRLSEVMQKVSQDDLIQELRNKLACSNAVLVLEKALIKLQA 282  
+ T +E +E Q+++ +K SQ + L ++ + +L LE ++LQ

Sbjct: 493 QEELTKKLQRE-REFQEQMKVALEK-SQSEYL--KISQEQEQLALEE---LELQK 544

Query: 283 DFASCTATHRYPPSSSECEEDIKKILKHLQEQKDSQCLHVEEYQNLVKDLRVELEV-SE 341  
A T + + +E + + L + + +E +N KDL V LEA ++

Sbjct: 545 K-AILTESENKLRLDQEAETRYTRILESSLEKS---LQENKQSKDLAVHLEAKNK 600

Query: 342 QKRNIHKDMKLELDLHGLREETSAAHIERKDKDITI-LQCRLOLEQLEFTEQKTLTKKD 400  
+ I + K + +L L+ + A K + + Q +++L+ E +K TL KD

Sbjct: 601 HNKEITVMVEKHKTELESILKHQDQDALWTEKLQVLKQYQTEMEKLR-EKCEQEKETLLK 659

Query: 401 K-----FLOEKDEM-LQELEKLTQVQNSLLKKEKELEKQCMATELMTVKEAKQDKS 453  
K ++E +E L++L+ K T+++ SL + E+ K + E +V + + DK

Sbjct: 660 KEIIFQAHIEEMNEKTLEKLDVKQTELE-SLSSELSEVLKARHKLEE-ELSVLKDQTKM 717

Query: 454 K-EAECKALQAEVQKLNKSLSEAKQERLAAQQAQC-KEEAALAGCHLEDTQRKLQKGL 511  
K E E K + + ++ ++ ++ Q+ + K++ L++ + L++

Sbjct: 718 KQELEAK-MDEQKNHHQQQVDSIIKEHEVSIQRTEKALKDQINQLELLKDKHKEHQ 776

Query: 512 L-LDKQKADTIQELQRELQMLQKSSMAEKEQTSNRKRVEELSLELSEALRKLNSDKK 570  
++ +AD I+ + ELQ + + + Q++ ++ + +L++ +KL + + E+

Sbjct: 777 AHVENLEAD-IKRSEGELOQASAKLDVFSYQSATHEQTKAYEEQLAQLQKLLDLETER 835

Query: 571 RQLQKTVAEQDMKMDM---LD--RIKHQHQSGSIK--CKLEEDLQEA TKLEDKREQL 623  
L K VAE + + D+ LD +I+ Q Q K ++E+ ++ T++ E K E

Sbjct: 836 ILLTKQVAEVEAQKQKDVCTELDAHKIQVDLMQLEKQNSEMEQKVKSLTQVYESKLEDG 895

Query: 624 KKSKEHEK--LMGELEALRQEFKKKOTLKENSRLKEENENLRAELQCCSTQLESSLN 681  
K +E K L+E E L+ +K K ++ ++KL + +++ + T+ ++

Sbjct: 896 NKEQEQTKQILVEKENMILQMRGQK-KEIEILTQKLSAKEDSIHILNEEYETKFRNQEK 954

Query: 682 KYNTSQOVIQDLNKEIALQKESLMSLQAOLDKALQKEKHYLQTTITKEAYDALSRKSAAC 741  
K +Q +++ + + K+ L+ +A+L K L E L+ + ++ ++A + A

Sbjct: 955 KMEKVKQKAKEMQETL---KKLLDQEA LKEL--ENTALELSQKEKQFNKMLEMAQA 1009

Query: 742 QD-DLTQALEKLNHVHTSETKSLQOSLTQTOEKKAQLEEEIIAYEERMKKLNTLRLGRGF 800  
++ A+ +L T++ + ++ SLT+ + +L + I +E KKLN + +L+

Sbjct: 1010 NSAGISDAVSRLE--TNQEQIE-SLTVHRR--ELNDVISIWE---KKLNQQAELQEI 1061

Query: 801 HQESELEVHAFDCKLEEMSCQVLQW--QKQHNDLKMMLAAKEEQLEFQEEAALKENLL 858  
H E+++ ++++ E+ ++L + +K+ N ++ KEE +++ + L+E L

Sbjct: 1062 H---EIQLEKEQEVAELKQKILLFGCEKEEMNK-EITWLKEEGVKQ-DTTINELQEQLK 1116

Query: 859 EDDKEPCCLPQWSPKDTCLRYRGNDQIMTNLEQ--WAKQKQVANEKLGQNLREQVNYI- 915  
+ L Q K L + + +L++ + ++Q V + L + + +V+ +  
Sbjct: 1117 QKSAHVNSLAQ-DETKLKAHLEKLEVDLNKSLKENTFLOEQQLVELKMLAEEDKRVSELT 1175

Query: 916 AKLSGEKDLHSMVMHLQOENKKLK-KEIEEKKMKAE 951  
+KL + S+ ++ NK L+ K +E KK+ E  
Sbjct: 1176 SKLKTDEEFQSLKSSHEKSNKSLKEDKSLFCKLSEE 1212

Score = 337 (50.6 bits), Expect = 4.0e-26, P = 4.0e-26  
Identities = 215/951 (22%), Positives = 433/951 (45%)

Query: 10 REVSSLSNKLQSLQDLKLNLDVCKRQKTLQDNQLCMEAMNSSHDKKQAQALAFEESE 69  
+E + +++L L+ ++ K Q K L + EA + H+K+ + E+ +  
Sbjct: 560 QEAETRYTRILELESSLEKSLQENKNQSKDLAVHL---EAEKNKHKNKEIT--VMVEKHK 613

Query: 70 VEPGSSKQCHLRQLQQLKKLLVLQOEFHTEELQTSYSLRQYQSIKELQTSDLVLH 129  
E S K H +Q +KL VL+Q+ + E+L+ Q + L K +++  
Sbjct: 614 TELESK--H-QQDALWTEKLVKQYQTEMEKREK---CEQEKETLLKD-KEIIFQA 666

Query: 130 HHCKLKE---DEVILYEEEMGNHNTGKEL---HLAQEQALAGDKIASLERSLNLYRD 183  
H + + E + + + + E + + + E L H +E+L+ + D+ + + + L D  
Sbjct: 667 HIEEMNEKTLEKLDVKQTELESLSSESEVLKARHKEELSVLKQDTQDKMKELEAKMD 726

Query: 184 K---YQSSLSNIELLECQVKMLQGE--LGGINGQEPENKGDHKSRIYTPSPCMIEHQE 237  
+ + +Q + + I + E + V + + E L + Q + K + + + + +  
Sbjct: 727 EQKNHQQQVDSI-IKEHEVSIQTEKALKDQINQLELLKERDK-HLKEHQAHVENLEA 784

Query: 238 TQKRLSEVMQKVSQDDLIQELRNKLACSNALVLEREKALIKLQADFASCTATHRYPPSS 297  
KR O+ S + D+ Q ++ ++ E+ L+LQ T R  
Sbjct: 785 DIKRSEGLQQAASAKLDVFSYQS---ATHEQTKAYEQLAQQLLDLE-TERIL--- 837

Query: 298 SEECEDIKKILKHLQEQKDSOCLHVEEYQNLVKDLRVLAEVSEQKRNIMKMMKL-ELD 356  
+ K + ++ OK C ++ ++ V+DL +LE + + +K + ++ E  
Sbjct: 838 -----LTKQVAEVEAQKDVCTELDAHKIQVDLMQOLEKQNSEMEQVKSITQVYESK 891

Query: 357 LH-GLREETSAPIERKDKDITILQCRIL-QELQLEFTEQTKLTKKDKF--LQEKDEM-LQ 411  
L G +E+ +K+ ILQ R Q+ ++E TQKL+ K+D L E+ E +  
Sbjct: 892 LEDGNKEQEQTKQILVEKENMILQMRGQKKEIIL-TQKLSAKEDSIHILNEEYETKFK 950

Query: 412 ELEKILTOVQNSLLK-----KEKELEKQCMATELMTVKEAKQDKSKEAECKALQAEVQ 466  
EKK+ +V+ + K+K L+++ + ELE T E Q K K+ K L+ Q  
Sbjct: 951 NQEKMKVQKAKEMQETLKKLLDQEAELKKELENTALELSQ-KEKQFNAKMLEM-AQ 1008

Query: 467 KLNLSLEAKQOERLAAQQAQCKEEAALAGCHLEDTRKRLQKGLLLDKQKADTIQELQR 526  
+A RL Q Q + + + L D +K L Q+A+ +QE+  
Sbjct: 1009 ANSAGISDAVS--RLETNQKEQIESLTVHRRELNDVISIWEKKL---NQAEELQEIH- 1062

Query: 527 ELQMLQKSSMAEKEQT-----SNRKR--EELSLESEALRKLNSDKKQRLQ 574  
E+O+ +KE +AE +Q K + +E ++ L +L+ K+K  
Sbjct: 1063 EIQLQKEQEVAELKQKILLFGCEKEEMKEITWLKEEGVKQDTTLNELQEQQLKQSAHV 1122

Query: 575 KTVAEQDMKMDMLDRIKHQHREQGSIKCKLEEDLQEAATKLEEDKREQLKKSKEHEKIME 634  
++A+ + K+ L++++ + L+E L E L E+ + + + K +  
Sbjct: 1123 NSLAQDETCLKAHLEKLEVDLNKSLKENTFLOEQQLVELKMLAEEDKRVSELTSLKLTDD 1182

Query: 635 GELEALRQEFKKDKTLKENSRLKEEENENLRAELQCCSTQLESSLNKYNTSQQVIQDLN 694  
E ++L+ +K +K+L++ S + ++ +E L +L C + E+ L T++ + +  
Sbjct: 1183 EEFQSLKSSHEKSNKSLKEDKSLFCKLSEELAIQLDICKKTEALLEA-KTNELINISS 1241

Query: 695 KEIALQKESLMSLQALDKALQKEKHYLQTTITKEAYDALSRKSAACQDDLT----QALE 750  
K A+ + Q + K KE ++T E +A R+ Q+ L QA  
Sbjct: 1242 KTNAILSR-ISHCOHRTTKV--KEALLIKTCTVSEL-EAQLRLTEEQNTLNSFQOATH 1297

Query: 751 KLNHVTSETKSLQSLTQTEKKAQLEEEIAYEERMKKLN---TELK--LRGFHQSE 805  
+L ++ KS++ + Q +K L++E ++ + T+L+K  
Sbjct: 1298 QLEEKENQIKSMKADIESLVTEKALQKEGGNQQAASEKESCITQLKKESENINAVTL 1357

Query: 806 LEVHAFDKKLE--EMSCQVLOWQKQKQNDLKMALAEKQLEFQEEAALKENLLEDDKE 863  
++ +KK+E +S Q+ Q QN + L+ KE + + + + K LL D +  
Sbjct: 1358 MKEELKEKKVEISSLSKQLTDLNVQLQNSIS-LSEKEAAISSLRKQYDEEKCELL-DQVQ 1415

Query: 864 PCCLPQWSPKDTCLRYRGNDQIMTNLEQNAKQKQVANEKLGQNLRE---QVNYIAKLSG 920  
++ K+ + D +W K+ + + N ++E Q+ +K +  
Sbjct: 1416 DLSFKVDTLSKEKISALEQVDMSNKFSEWKKKAQSRFTQHONTVKELOIQLKSKSEAY 1475

Query: 921 EKDH-LHSMVMHLQOENKK---LKKEIEEKKMKAE 951  
EKD ++ + L Q+NK+ LK E+E+ K K E  
Sbjct: 1476 EKDEQINLLKEELDQONKRFDCLEKGEEDDKSKME 1510

Score = 332 (49.8 bits), Expect = 1.4e-25, P = 1.4e-25  
Identities = 209/953 (21%), Positives = 438/953 (45%)

Query: 1 MKDEAGERDREVSSLSKLLSLQLDIKNLHDVCKRQRKTLDQNLQCMEEAMNS----SHD 56  
 Sbjct: 470 MKKSSEEQIAKLQKLHEKELARK-EQELTKKLTREREFEQMKVALEKSQSEYLKISQE 528

Query: 57 KKQAQALAFEESEVEFGSSKQCHLRQLQQLKKLLVLQOELEFTEELQTSYSLRQYQS 116  
 Sbjct: 529 KEQESLALAELELQ---KKAILTESEN---KLRDLQOEAEYTRILELESSLEKSQ 581

Query: 117 ILEKQTSDLVLLHHHCKLKEDE--VILYEE----EMGNHNENT--GEKLHLAQEQLALA 167  
 Sbjct: 582 ENKNQSKDLAVHLEAEKNKHNKEITVMVEKHKTELESKHOQDALWTEKLQVLKQYQTE 641

Query: 168 GDKIASL--ERSLNLYRDK---YQSSLS--NIELLECQVKMLQOGELGGINGQEPENKGDH 220  
 Sbjct: 642 MEKLREKCEQEKETLLKDKIIFQAHIEEMNEKTLE-KLDVKQTELESLSSELSEVLKAR 700

Query: 221 SKVRIYTPCMIQEHQETQKRLSEVMQKVSQDDLIQELRNKLACSNALVLEKALIKL 280  
 Sbjct: 701 HKLEELS--VLKD--QTDKMKQLEAKMDEQKNHQQQVDSIIKEHVSIRTEKALKD 756

Query: 281 QADFASCTATHR--YPPSSSEECEDIKKILKHLQEQKDSQCLHVEEYQNLVKDLRVELEA 338  
 Sbjct: 757 QINQLELLKERDKHLKEHQAHVENLEADIKRSEGLQQAASAKLDVFSYQSATHEQTKA 816

Query: 339 VSEQKRNIMKDMKLELDLHGLREETSABIERKOKDITILQCRLOLEQLETFETQKLT 398  
 Sbjct: 817 YEEQLAQLQKLLDLETERILLTKQV-AEVEAKKDV---CT--ELDAHKIQVQDLMQQ 869

Query: 399 KDKFLQEKDEMLQLEKLLTQVQNSLLK-EKELEKQOCMATELEMTVKEAKQDKSEAE 457  
 Sbjct: 870 LEK---QNSEMEQKV-KSLTQVYESKLEDGKQEQOTKQILVEKENMILQMGREGQKKEIE 925

Query: 458 C--KALQAEVQKLKNSLEEAQKQERLAAQQAQCKEEAALAGCHLEDTRK--LQKGLL 513  
 Sbjct: 926 ILTQKLSAKEDSIHILNEEYETKFKNQEKMEKVKQKAK---EMQETLKKLLDQEA 981

Query: 514 DKQKADTIQEL-QRELQMLQKSSMAEKQTSNRKRVEELSLESEALRKLNSDKERKQ 572  
 Sbjct: 982 KKELENTALELSQKEQFNAKMLEMAQANSAGISDAVSRLETNQEIESL--TEVHRRE 1039

Query: 573 LQKTVAEQDKMMNDLRIKHQHREQGSIRCKLEEDLOEATKLEEDKREOLKKS---KE 628  
 Sbjct: 1040 LNDVISIWEKKLNQAELEQEIHEIQLEKEQEVAEKQKILLFGCEKEEMKEITWLKE 1099

Query: 629 HEKLMEGEALRQEFKKDKTLKENSRLKEENENLRAELQCCSTOLESSLNKYNTSQ 688  
 Sbjct: 1100 EGVKQDTTLNELQEQKSAHV--NS--LAQDETCLKAHLEKLEVLNLSKLENTFLQE 1155

Query: 689 VIQDLNKEIALQKESLMSLQAQL---DKALQ--KEKHYLOTTITKEA---YDALSRSKAA 740  
 Sbjct: 1156 QLVELKMLAEEDKRVSELTSLKTTDEEFQSLKSSHEKSNSKLEDSLEFKKLSSE-LA 1214

Query: 741 CQDDL----TQAL-----EKLNVHTSETKSLQOSLTQTQEKKAQLEEEIIAYEERMKKL 790  
 Sbjct: 1215 QLDICCKTEALLEAKTNELINISSKTNAILSRISHCQHRRTKVKEALLIKTCTVSEL 1274

Query: 791 NTELRLRGFHWQSELEVHAFDKKLEEMSCQVLOWQKHQNDLKMLAAKEEQLEFQEM 850  
 Sbjct: 1275 EAQLROLTEEQNTLNISFQATHQLEKENQI---KSMKADIESLVTEKEAL---QREG 1327

Query: 851 AALKENLLEDDKEPCCLPQWSVPKDTCLRLYRGNDQIMTNLEQWAKQKQVANKEKLGOLAE 910  
 Sbjct: 1328 G--NQQAASEKESC-ITQ--LKKELSE---NINAVTLMKEELKEKKVEISSLSKQLTD 1378

Query: 911 ---QVNYIAKLSGEKDLHLSVMVHLQENKLLKEIEEKKMAE 951  
 Sbjct: 1379 LNVQLQNSISLSEKAAISSLRKQYDEEKCELLDQVQDLSFKVD 1422

Score = 329 (49.4 bits), Expect = 2.9e-25, P = 2.9e-25  
 Identities = 226/941 (24%), Positives = 444/941 (47%)

Query: 61 QALAFEESEVE--FGSSKQCHLRQLQQLKKLLVLQOELEFTEELQTSYSLROYOSIL 118  
 Sbjct: 165 QMLQREKKLQGISQSDKSLRRIAELEELQMDQAKHLQEEFDASLEEKQYISVL 224

Query: 119 EKQTSDLVLLHHHCKLKEDEV-----ILYEEEMGNHNENT--GEKL---HLAQEQLALA 167  
 Sbjct: 225 QTVQSLRLRNGPMNVDLKPLQLEPQAEVFTKEENPESDGEFVVDGTSVKLTETL 284

Query: 168 GDKIASLERSLNLYRDKYQSSLSNIELLECQVKMLQOGELGGINGQEPENKGDHDKVRIYT 227

Sbjct: 285 QQRVQRQENLLKRCKETIQSHKEQCTLLTSEKEALQEQLDERL-QELEKIKD---LHMAE 340  
 Query: 228 SPCMIQEHQETQKRLSEVWQKVSQDDLIQELRNKLACSNAVLEREKALIKLQADFASC 287  
 +I + ++ + ++ Q +I E + ++ L ++ E+ + +L++  
 Sbjct: 341 KTKLITQLRDAKNLIEQLEQDKGM---VIAETKRQM--HETLEMKEEE-IAQLRSRIKQM 394  
 Query: 288 TATH---RYPPSSSEEC--EDIKKILKHLQEQKDSQCLHVEEYQNLVKDL-----RVE 335  
 T R SE E+++K L Q+ +++++ E +K + R+  
 Sbjct: 395 TTQGEELREQEKSERAFAEELEKALSTAQKTEARRKKAEMDEQIKTIEKTSEERIS 454  
 Query: 336 LEA-VSEQKRNIMKDMML--ELDLHGLREETSABIERKDKDITILQCRLOLEQLETTET 392  
 L+ +S K+ ++ D+MK E + L++ + RK++++T +LQ + EF E  
 Sbjct: 455 LQQLSRVQEVV-DVMKSSSEQIAKLQKLHEKELARKEQELTK---KLQTREREFQEQ 510  
 Query: 393 QKLTLLKKDKFLQEKDEMLQLEKKLTQVQNSLLKKEKELEKQCCMATELMTVKEAKQDK 452  
 K+ L+K + E ++ QE E+ Q SL +E EL+K+ + TE E +++ +Q+  
 Sbjct: 511 MKVALEKSQ--SEYLKISQEKQ-----QESLAELEELQKKAIL-TESENKLRDLQEQ- 561  
 Query: 453 SKEAECKALQAEVQKLNKSLLEAKQOER-----LAAQQAQCKEEAALAGCHLEDTOR-K 506  
 ++ + L+ E L+ SL+E K Q + L ++ KE + H + + K  
 Sbjct: 562 AETRYRTRILELE-SSLEKSLQENKNSKDLAVHLEAKNKHNEITVMVKEKHKTELESK 620  
 Query: 507 LQKGLLLDKQADTIQELQREQLQKESMAEKEQTSNRKRVEELSLELSEALRK-LEN 565  
 Q+ L ++ Q+ Q E++ L +E EKE K + + E K LE  
 Sbjct: 621 HQQDALWTEKLQVLKQOYQTEMEKL-REKCEQEKETLLKDKKII-FOAHIEEMNEKTLEK 678  
 Query: 566 SDKEKRLQKQTVAEQDNKMNMDLRIKHQHQEGSI-KCKLEEDLQEA-TKLEDDKR--E 621  
 D ++ +L+ +E ++++L + +H+ E+ S+ K + ++ QE K+ E K +  
 Sbjct: 679 LDVKQTELESLSSE---LSEVL-KARKMLEELSULKDQTDKMKQLEAKMDEQKNHHQ 733  
 Query: 622 QLKKS--KEHEKLMGELEALRQEFKKDKDTLKENSRLKEEN---ENLRAELQCCSTQL 676  
 Q S KEHE ++ +AL+ + + + LKE + L+E ENL A+++ +L  
 Sbjct: 734 QQVDSIIKEHEVSIQRTKALKDQINQLELLKDKDKLKEHQAHVENLEADIKRSEGE 793  
 Query: 677 ESSLNKYNTSQVQIDLNKEALQKESLMSLQAQLDKALQKEKHLYLTTITKEAYDALSR 736  
 + + K + Q +++ +E L LQ +L L+ E+ L TK+ + ++  
 Sbjct: 794 QQASAKLDVFQSYQSATHEQTKAYEEQALQOQL-LDLETERILL---TKQVAEVEAQ 848  
 Query: 737 KSAACQD-----DLTQALEKLNHVTSKSLQOQSLTQTOEKKQ--LEEEIAYEE 785  
 K C + DL Q LEK N SE + +SLTQ E K + +E+ +  
 Sbjct: 849 KKDVCTELDAHKIQVQDLMOQLEKQ--SEMEQKVKSLTQVYESKLEDGKKEQEQTKQI 905  
 Query: 786 RMKLNTELRKRGFHOESELVHAFDKKLEEMSCQVL--QWQKHQNDLKLMAAKEEQ 843  
 ++K N L+ G Q+ E+E+ +E S +L +++ +N K + ++  
 Sbjct: 906 LVEKENMILQMRG--QKKEIILTQKLSAKEDSIHILNEEYETKFKNQEKMEKVKQKA 963  
 Query: 844 RFQOEEMAALKENLEDDKEPCCLPQWSVPKDTCLRYRGNDQIMTNLEQWAKQKQV--- 899  
 +E QE LK+ LL+ + + L + + L + Q + + A+  
 Sbjct: 964 KEMQE---TLKKLLDQEA---LKK-ELENTALELSQKEKOFNAKMLEMAQANSAGISD 1016  
 Query: 900 ANEKLGNLREQVNYIAKLSG-EKDHLHVMVH-LQENKKLKK--EIEEKKMAENTRL 955  
 A +L +EQ+ + ++ E + + S+ L Q+ ++L++ EIT+ ++ + E L  
 Sbjct: 1017 AVSRLETNQKEQIESLTVHRRELNDVISIWEKKLNQQAEELOEIHEIQLEKEQEVAEL 1076  
 Query: 956 CTKALGPSRTSTQREKVCGLGWKGLPQD 985  
 K L E + K L +G+ QD  
 Sbjct: 1077 KQKIL-LFGCEKEEMNKETLWKEEGVKQD 1105  
 Score = 326 (48.9 bits), Expect = 6.0e-25, P = 6.0e-25  
 Identities = 220/907 (24%), Positives = 444/907 (48%)  
 Query: 67 ESEVEFGSSKQCHLRLOQLKKLLVLQOEFHTEELQTSYSLRQYSILE---KOTS 123  
 E+E G+S + QL Q +++ EL T+Y L++ + L+ Q+  
 Sbjct: 123 EAEDLVGNSDSLNEQLIQLRRMERSLSSYAGKYSVLTAYQMLQREKKLQGISLSQ 182  
 Query: 124 DLVLLHHKCKEDEVILYEEEMGNHNTGEKHLHAEQLALAGDKIASLERSLNLYRD 183  
 D L +L+E+ + +++ H + E+ + E+ I+ L+ +L+ +  
 Sbjct: 183 DKSL-RRIAELREE--LQMDQAKKHLQ---EEDFASLEE---KDQYISVLQTVSLLKQ 233  
 Query: 184 KYQSSLSNIELLECQVKMLQELGGINGQE-PENKG-----DHKSVR-IYTPCMIQEHQ 236  
 + ++ N+++L+ + L+ + +E PE+ G D + V+ + T ++ +  
 Sbjct: 234 RLNGPMNVDLK-PLPQLEPQAEVFTKEENPESDGEPPVEDGTSVKTLETIQQRVKQE 292  
 Query: 237 ETQKRLSEVWQKVSQDDLIQELRNKLACSNAVLEREKALIKLQADFASCTATHYPPS 296  
 ER E Q +Q L+ K A L ER + L K+ D T  
 Sbjct: 293 NLLKRCKETIQSHKEQCTLLTS--EKEALQEQLD-ERLOLEKIK-DLHMAEKTTLIT-- 346  
 Query: 297 SSEECEDIKKILKHLQEQKDSQCLHVEEYQNLVKDLRVELEAVSEQKRNIMKDMMLELD 356  
 + DK +++ L++ K + E + + + L ++ E ++ Q R+ +K M +  
 Sbjct: 347 ---QLRDAKNLIEQLEQDKGM---VIAETKRQMHEHTEMKEEEIA-QLRSRIKQMTTQGE 400

670

Sbjct: 1783 AE K Q + + L + LEE ++ L Q + + + A +LE+ +QK L  
Query: 512 LLDKQKADTIQELQRELOMLQKESMAEKEQTSNRKRVEELS--LELSEALRKLENSDKE 569  
++K T Q L++++ L +S + +++ +R +EEL+ E +AL++++ +K  
Sbjct: 1843 --QEKELTCQILEQKIKEL--DSCLVRQKEV-HRVEEELTSRYEKQLQALQOMDGRNKP 1896  
Query: 570 KRQLOKTVAEQD---MKMNDMLDRIKHQHREQSGIKCKLEEDLQEATKLLDKREQLKK- 625  
L++ E+ + +L ++ QH + E + Q+ K + ++ L+  
Sbjct: 1897 TELLEENTEEKSKSHLVQPKLLSNMEAQHNDLEFLAGAEERKQLGKEIVRLQKDLRML 1956  
Query: 626 SKEHEKLMGELEALRQEFKKKDKTLKENSRLKEENENLRAELQCCSTQLESSLNKYNT 685  
KEH++ ELE L++E+ + E K+++E E+L EL+ ST L+ + ++NT  
Sbjct: 1957 RKEHQ--EELILKEYDQ-----EREKIKQEDEL--ELKHNT-LQLMREFNT 2003  
Query: 686 S-QQVIQDLNKEIALQKESLMSLQAQLDKALQKEKHYLQTTITKEAYDALSRKSAACQDD 744  
Q Q+L I ++A+L ++ Q+E + L I E D L R +A ++  
Sbjct: 2004 QLAQKEQELMTIKETINKAQEVEAELLESHEETNQLLKIA-EXDDDLKR-TAKRYEE 2061  
Query: 745 LTOALEKLNHVTSKSLQOQSLTQTEKKAQ-LEEELIAYEERMK--KLNTLRLKRGFH 801  
+ A E+ +T++ + LQ L + Q+K Q LE+E + + +L T+L +  
Sbjct: 2062 ILDAREE--ENTAKVRDLQTOLEELQKKYQQKLEQENPGNDNVTIMELQTLAQKTTLI 2119  
Query: 802 QESELEVHAFDKKLEEMSCQVLOWQK 827  
+S+L+ F +++ + ++ ++K  
Sbjct: 2120 SDSKLKEQEFREQIHNLEDRLLKYEK 2145  
Score = 316 (47.4 bits), Expect = 7.1e-24, P = 7.1e-24  
Identities = 213/977 (21%), Positives = 454/977 (46%)  
Query: 4 EAGERD-REVSSLSKLLSLQLD-INKLHDVCKRQRKTLQDNQLCMEAMNSSHOKKQAO 61  
E R+ +V S+ K L+ Q + + +H++ + Q K + +L + + ++ +  
Sbjct: 1034 EVHRRELNDVISIWEKKLNQQAELQEIHEI-QLQKEQEVaelKQKILLFGCEKEEMNK 1092  
Query: 62 ALAFEESEVEFGSSKQCHLRQLQ-QLKKLL---VLQOE--LEFTEELQTSYSLRQY 114  
+ + + E G + L +LQ QUK+K + Q E L+ H E L+ +  
Sbjct: 1093 EITWLKE--GVKQDTTLNELQELQKQSAHVNSLAQDETCLKAHLEKLEVDLNSLKE 1149  
Query: 115 QSILEKQTSDLVLLHHHCKLEDEV---ILYEEEMGNHNTGKGLHLAQELALAGDKI 171  
+ L++Q +L +L K K E+ + +E ++ EK + +E +L K+  
Sbjct: 1150 NTFLEQVLVFLMLAEEDKRRVSELTSLKLTDEEFQSLKSSHEKSNKSLKEDKSLEFKKL 1209  
Query: 172 AS-LERSLNLVYRDYQSSLS--NIELLECQVKMLQELGGINGQEPENKGOHSKVRITYS 228  
+ L L++ K ++ L EL+ L I +++ K +  
Sbjct: 1210 SEELAIQLDICCKTEALLEAKTNELINISSKTNAILSR--SHCQRTTKVKEALLIK 1267  
Query: 229 PCMIQEHQ-----ETQKRLSEVQKVSQO-DDLIQELRNKLACSNAVLREKALIKL 280  
C + E + E Q L+ +Q+ + Q ++ +++ A +LV E+E L  
Sbjct: 1268 TCTVSELAQLRQLTEEQNTLNISFQQAHLQEEKNQIKSMKADIESLVTEKEA---L 1323  
Query: 281 QADFASCTATHRYPPSSSEECEDIKKILKHLQEQKDSQCLHVEEYQNLVKDLRVELEAVS 340  
Q + + + S E C I ++ K L E ++ L EE +K+ +VE+ ++S  
Sbjct: 1324 QKEGNN---OQQAASEKESC--ITQLKKESENINAVTLMKEE---LKEKKVEISSLS 1373  
Query: 341 EQKRNIMKDMKLELDLHGLREETSABIERKDKDITILQCRLOEL--QLEFTETQKLT-L 397  
+Q ++ + + L S+ ++ D++ L ++Q+L +++ +K++ L  
Sbjct: 1374 KQLTDLNVQLQN-SISLSEKAAISSLRKQYDEEKCELLDQVQDLSFKVDTLSKERISAL 1432  
Query: 398 KK-DKFLQEKDEMLQLEKKLTQVQNSLLKKEKELEKQOCMATELEMTV---KEAQDKS 453  
++ D + + E ++ + + TQ QN++ + +LE + A E + + KE ++  
Sbjct: 1433 EQVDDWSNKFSEWKKKAQSRFTQHQNTVKELQIQLELKSKEAYEKDEQINLLKEELDQQN 1492  
Query: 454 KEAECKALQAEVQKLKNSLEEAQQRERLAAQQAQKEEAALAGCHLE-DTORKLQKGLL 512  
K +C + E K K +E+ + L +Q A + E + +E ++ ++K  
Sbjct: 1493 KRFDCLEKGEEDDKSMKESNLETELKSTARIMELEDHITQKTIEIESLNEVLKNY- 1551  
Query: 513 LDKQKADTIQELQRELOMLQKESMAEKEQTSNRKRVEELSLELSEALRKLENSDKEKQ 572  
+ +QK +EL ++LQ Q+ + +++ L ++ +LE KE  
Sbjct: 1552 -NQKDIHKELVQLQHFQELGEEKDNVKEAEKILTLENQVYSMAKAELETKKLEH 1610  
Query: 573 LQKTVAEQDMKNDMLDRIKHQHREQ-GSIKCKLEEDLQEATKLL---EDKREQLKKS 627  
+ +V ++ ++ + DR++ + + +K K E+ + K L E+K EQ KK  
Sbjct: 1611 VNLVSKSKEELKALEDRLESESAKLAELKRAEQIAAIKKQLLSQMEKEEQYKGT 1670  
Query: 628 EHEKLMGELEALRQEFKKKDKTLKENSRLKEE-ENENL---RAELQCCSTQLESSLNK 682  
E EL QE +++ L+E + +E ++E L A+ T+ E + ++  
Sbjct: 1671 ESHL---SELNTKQEREREVHILEEKLKSVSSQSETLIVPSAKNVAAYTEQEADSQ 1727  
Query: 683 ---YNTSQQVIQDLNKEIALQKESLMSLQAQLDKALQKEKHYLQTTITKEAYDALSRKSA 739  
T ++ I L + + +KE L+ Q +K H+ +E L A  
Sbjct: 1728 GCVQKTYEEKISVLQRNLT-EKEKLLQRVGQ-EKEETVSSHFMRCQYQERLIKLEHAE 1785

Query: 740 ACQDDLTQALEKLNHVTSET--KSLQSSLTQTQEKKAQLEEEIIAYEERMKKLNTLRLK 797  
 +D Q++ +H+ E K++ SL Q + + + I ++ ++ + ++K  
 Sbjct: 1786 KQHED--QSM--IGHLQEELEKNKKYSLIVAQHVKEGGKNNIQAQNLNENFDDVQRT 1841

Query: 798 RGFHQESELVHAFDKKLEEM-SCQVLQWQKHQNDLKMLAAKEEQLEFQEEMAALKEN 856  
 QE EL ++K++E+ SC V Q ++ H+ +++ L +K E+L+ Q+ K  
 Sbjct: 1842 L---QEKELTCQILEQKIKELDSCLVRQ-KEVHRVEMEELTSYKELQALQMDGRNKPT 1897

Query: 857 -LLEDDKEPCCLPQWSPVKDTCRLYRGNDQIMTNLEQWAKQKQVANKEKLGNLREQVNYI 915  
 LLE++ E PK + ++ + L A++K +KLG ++ +  
 Sbjct: 1898 ELLEENTEESKSHLVQPKLLSNMEAQHNDFKLAG-AEREK---QKLGKEIVRLQKDL 1953

Query: 916 AKLSGE-KDHLHSMVMVHLQENK-KLKEIEEKKMAENTRLCTKALGPSRTESTOREK 972  
 L E + L + QE + K+K+E + ++K +T + + + T+ Q+E+  
 Sbjct: 1954 RMLRKEHQEELILKKEYDQEREKIKQEDELKHNST--LKQLMREFNTQLAQREQ 2010

Score = 301 (45.2 bits), Expect = 2.9e-22, P = 2.9e-22  
 Identities = 221/952 (23%), Positives = 441/952 (46%)

Query: 1 MKDEAGERDREVSSLSKLLSLQLDIKNLHDVCKRORKTLDQNOL---CMEEAMNSSHD- 56  
 +K A E R+VS L SKL + ++L ++ K+L+D L + E + D  
 Sbjct: 1160 LKMLAEEDKRVSELTSLKTTDEEFQSLKSSHEKSNKSLDKSLEFKLSEELAIQLDI 1219

Query: 57 --KKQAQALAFESE-VEFGSSK-QCHLRQLQQLKKLLVLQOELEFHT---EELQTSYY 109  
 KK L + +E + SSK L ++ + + ++ L T EL+  
 Sbjct: 1220 CCKKTEALLEAKTNELINISSKTNAILSRISHCHQRTTKVKEALLIKTCTVSELAQLR 1279

Query: 110 SLRQYQSILEKQTSQDLVLLHHCKLKEDEVILYEEEMGNHNTGERLHLAQE---QLAL 166  
 L + Q+ L H + KE+++ + ++ EK L +E Q  
 Sbjct: 1280 QLTEEQNTLNISFQAT---HQLEEKENQIKSMKADI---ESLVTEKEALQKEGGNQQA 1333

Query: 167 AGDKIASLERSLNLYRDKYQSSLSNIELLECQVRLQGEELGGIMQEPENKGDHVKVRIY 226  
 A +K E + + + ++ + L++ +K + E + Q + V++  
 Sbjct: 1334 ASEK---ESCITQLKSELSENINAVTLKKEELKEKKVEISSLSKLTD----LNVQLQ 1384

Query: 227 TSPCMIQEHQETQKRLSEVWQKVSQDDLIQELRNKLACSNALVLREKALIKLQADFAS 286  
 S + ++ + + + + D +Q+L K+ + L E+ AL + + D+++  
 Sbjct: 1385 NSISLSEKEAAISSLRKQYDEEKCELLDQVQDLSEKVV---DTLSKEKISALEQVD-DWSN 1440

Query: 287 CTATHRYPPSS--SEPECIDIKKHLKHLQEQKDSQCLHVEEYQNLVKD-----LRVE-LE 337  
 + + S ++ +K++ L E K + +E NL+K+ R + L+  
 Sbjct: 1441 KFSWKKAQSRFTQHONTVKELQIQL-ELKSKEAYEKDEQINLLKEELDQKNRFDCLK 1499

Query: 338 AVSEQKRNIM-KMMKLELDLHGLRE---ETSAHIERKDKDITILQCRQLQEL-QLETFET 392  
 E ++ M K LE +L E HI +K +I L L+ Q + E  
 Sbjct: 1500 GEMEDDKSKMEKKESNLETLSQTARIMELEDHITQKTIEESLNEVLKNYNQOKDIEH 1559

Query: 393 QKLTLLKDKFLQ---EKDEMLQELKLLTQVQNSLLKKEKELEKQOCMALENTVKEAK 449  
 ++L K F + EKD ++E E+K+ ++N + + ELE ++ + ++VK  
 Sbjct: 1560 KELVQKLQHFQELGEEKDNVKEAEEKILTLENQVYSMAELETKKKELEHVNLSVK--- 1616

Query: 450 QDKSKEAECKALQAEVQKLKNSLEEAKQOERLAAQQAQCKEEAALAGCHLEDTORLQK 509  
 SKE E KAL+ ++ S + + +R A Q+ A K++ +E+ + + +K  
 Sbjct: 1617 ---SKEELKALEDRLES--ESAALAEKRAEQIAAIAIKQLL---SQMEEKEEQYK 1668

Query: 510 GLLLDKQKADT-IQELQRELQMLQKESMAEKEQTSNRKRVESLSLESEALRKLENSDK 568  
 + + +T +QE +RE+ +L+++ E Q+ + S + A + E +D  
 Sbjct: 1669 GTESHLSELNKLQEREREVHILEEKLKSVESSESQSETL--IVPRSAKNVAAYTEQEADS 1726

Query: 569 E---KRLQK-TVAEQDMKMD-MLDRIKHQHQSGSIKCKLEEDLQEATKLLDKREQ 622  
 + K +K +V +++ + +L R+ Q +E ++ + E Q +L+ K E  
 Sbjct: 1727 QGCVQKTYEEKISVLQRLNTEKEKLLQVRG-QEKEE-TVSSHFMRCQYQERLI--KLEH 1782

Query: 623 LKKSKEHE-KLMGEL-EALRQEFKKKDKTKENSRLKEEENENLRAELQCCSTQLESSL 680  
 + +K+HE + M G L E L :+ KK + ++ K E N++A+ LE  
 Sbjct: 1783 AE-AQKHEDQSMIGHLOEELEKNKKYSLIVAQHVKE-EGGKNNIQAQ-----QNLE--- 1832

Query: 681 NKYNYSQQVIQDLNKEITALQKESLMSLQAQLDKAL--QKEKHVLOTITKEAYDALSR-K 737  
 N ++ Q+ +Q+ KE+ Q L +LD L OKE H ++ Y+ L +  
 Sbjct: 1833 NVFDDVQKTLQE--KELTCQ--ILEQKIKELDSCLVRQKEVHRVEMEELTSYKELQALQ 1888

Query: 738 SAACQDDLTQALEKLNHVTSETKSLQSSLTQTQEKKAQ--LEEEIIAYEERMKKLNTLRL-- 794  
 ++ T+ LE+ S++ +Q L E + LE ++ E +KL E+  
 Sbjct: 1889 QMDGRNKPTLELEENTEESKSHLVQPKLLSNMEAQHNDFKLAGAEREKQKLGKEIVR 1948

Query: 795 --RKLGRGFHQESELVHAFDKKLEEMSCQVLQWQKHQNDLKMLAAKEEQLEFQEEMA 852  
 + LR +E + E+ K+ ++ + ++ Q+Q +LK + ++ +REF ++A  
 Sbjct: 1949 LQKDLRLRKEHQEELILKKEYDQEREKIK-QEDELKHNSTLKQLMREFNTQLAQ 2007

Query: 853 LKENLLEDDKEPCCLPQWSPVKDTCRLYRGNDQIMTNLEQWAKQKQVANKEKLGNLREQV 912



Sbjct: 2008 ++ L KE Q V + + Q TN Q K K+A EK + R  
KEQLEMTIKETINKAQ-EVEAEELLES----QEETN--QLLK--KIA-EKDDDLKRTAK 2057  
Query: 913 NYIAKLSGKDHLSVMVHLQOENKKLKEIEKKMKAEN 952  
Y L ++ + + LQ + ++L+K+ ++K + EN  
Sbjct: 2058 RYEEILDAREEEMTAKVRDLQTQLEELQKKYQKLEQEN 2097

Score = 300 (45.0 bits), Expect = 3.7e-22, P = 3.7e-22  
Identities = 195/961 (20%), Positives = 435/961 (45%)

Query: 1 MKDEAGERDREVSSLSKLLSLQDIKN--LHDVCKRQRKTLQDNQLCMEEAMNSSHDKK 58  
+KD+ + +N K L +LD+K L + + L+ +EE ++ D+  
Sbjct: 657 LKDKIEIFQAHIEEMNEKTLE-KLDVKQTELESLSSELSEVLKARHK-LEEELSVLKQQT 714  
Query: 59 QAQALAFESEVEFGSSKQCHLRQLQKLLV-LQEEFHTTEELQTSYSLRQYQSI 117  
+E E + K H +Q+ +K+ V +Q+ + +++ L++  
Sbjct: 715 DKMK--QELEAKMDEQKNHQQQVDSIIKEHEVSIQRTKALKDQINQLELLKERDKH 771  
Query: 118 LEKQTSDLVLLHHCKLKEDEVILYEEEMG--NHNENTGEKHLAQEQLALAGDKIASL 174  
L++ + + L K E E+ ++ T E+ +EOLA K+ L  
Sbjct: 772 LKEHQAHVENLEADIKRSEGELQQAASAKLDVFSYQSATHEQTKAYEEQLAQQLKLLDL 831  
Query: 175 ERSNLNRYDKYQSSLSNIELLECQVKMLQEGELGIMGO-EPENKGDHKSRIYTSPCMIQ 233  
E L + + + + ++ +M Q E +N KV+ T  
Sbjct: 832 ETERILLTKQVAEVAQKQDVCTELDAHKIQVDLMQQLKQNSEMEQKVKSLTQ-VYES 890  
Query: 234 EHQETQKRLSEVWQKVSQDDLIQELRN---KLACSNALVLREKALIKLQADFASCTA 289  
+ ++ K + Q + ++++I ++R ++ + +E ++ L ++ +  
Sbjct: 891 KLEDGNKEQEQTQKILVEKENMILQREGQKKEIEILTQKLSAKEDSIHILNEEYET--- 947  
Query: 290 THRYPPSSSEECEDIKKILKHLQEQKDSQCLHVEEYQNLVKDLRVELEAVSEQKRNIMKD 349  
++ + ++ E +K+ K +QE + L E L K+L +S++++  
Sbjct: 948 --KFK-NOEKKMEKVKQKAKEMQETLKKLLDQEA--KLKKELENTALELSQEKQFNAK 1002  
Query: 350 MMKL-ELDLHGLREETA-HIERKDKDITILQCRLEQLQLETTETQKTLKKDKFLQEKD 407  
M+++ + + G+ + S +K++ ++ +EL + +K ++ +LQE  
Sbjct: 1003 MLEMAQANSAGISDAVSRLETNQKEQIESLTVHRRLENDVISIMEKKNLQAAELQEIH 1062  
Query: 408 EM-LQELEKRLTQVQNSLLK---KEKELEKQCMATE----LEMTVKEAQD-KSKEAC 458  
E+ LQE E+++ ++ +L +++E+ K+ E +T+ E ++ K K A  
Sbjct: 1063 EIQLEKEQEVAELKQIILLGCEKEEMKEITWLKEEGVKQDTTLNELQQLKQSAHV 1122  
Query: 459 KALQAEVQKLSLEAKQERLAAQQAQCKEAAAGCHLEDTQKRLQKGLLDKQKA 518  
+L + K LK LE+ + +E+ + + +E+ +RK+ L K K  
Sbjct: 1123 NSLAQDETCLKAHLKLEVDLNLKSLKENTFLQEQVLKMLAEEDKRRVSE--LTSKLT 1180  
Query: 519 DTIQELQRELQMLQESSMAEKEQTSNRKRVEELSLELSEALRKLNSDKERQLQKTV 578  
T+E Q +K + E + +K EEL+++L +K E +K + +  
Sbjct: 1181 -TDEEFQSLKSSHEKSNKSLKEDSLFKKLSLELAQLDICKKTEALLEAKTN--ELIN 1237  
Query: 579 EQDMKMNMDLRKH-QHREQGSICKLEEDLQEAATKLEDKREQLKSKSEHEKLMGEL 637  
K N +L RI H QHR K++E L T + + QL++ E + +  
Sbjct: 1238 ISSKTNAILSRISHCOHRTT----KVKEALLIKTCTVSELEAQLRQLTEEQNTLNISF 1292  
Query: 638 EALRQEFKKD---KTLKENSRLKEENENLR-----AELQCCSTQLESSL---- 680  
+ + ++K+ K+K + L E E L+ +E + C TQL+ L  
Sbjct: 1293 QQATHQLEKENQIKSMKADIESLVTEKEALQKEGNGQQAASEKESCITQLKKELSENI 1352  
Query: 681 NKYNTSQVQIDLNKEIALQKESLMSLQAQLDKALQ-KEKHYLQTTITKEAYDALSRKSA 739  
N ++ +++ EI+ + L L QL ++ EK +++ K+ YD +  
Sbjct: 1353 NAVTIMKEELKEKKVEISSKQLTDLNVQLQNSISLSEKAAISSLRKQ-YDEEKCELL 1411  
Query: 740 ACQDDLTQALEKLN-HVTSETKSLQSSLTQTQEKKAQLEEEIAYEERMKKLNTLRL-KL 797  
DL+ ++ L+ S + + + E K + + +K+L +L K  
Sbjct: 1412 DQVQDLSFKVDLTSKEKISALEQVDDWSNKFSEWKKKAQSRFTQHONTVKELQIQLELS 1471  
Query: 798 RGFHQESELEVHAFOKKLEEMSCQVLQWQKHQNDKMLAAKEEQLR-EFOEEMAALKEN 856  
+ +++ E +++ ++L+++ + + ++D + KE L E + + A + E  
Sbjct: 1472 KEAYEKDE-QINLLKELDQQNKRFDCLEGMEDDKSKMEKKESNLETLSQTARIME- 1529  
Query: 857 LLEDDKEPCCLPQWSPKDTCLRYRGNDQIMTNLEQWAKQKQVANEKLGNLREQVNYIA 916  
LED + + T + N+ ++ N Q OK K +L +++ +  
Sbjct: 1530 -LEDH-----ITQRTIEIESLNE-VLKNNYQ----QKDIENK---ELVQKLQHFQ 1570  
Query: 917 KLSGEXDH----LHSMVHLQOENKKLKEIEKKMKAENTRLCTKA 959  
+L EKD+ ++ L+ + +K E+K K + E+ L K+  
Sbjct: 1571 ELGEEKDNRVKAEAEKILTLENQVYSMAKAELETKKKELEHVNLSVKS 1617

Score = 298 (44.7 bits), Expect = 6.1e-22, P = 6.1e-22  
Identities = 207/886 (23%), Positives = 412/886 (46%)

Query: 47 MEEAMSSHDKKQAALAFEESEVEFGSSKQCHLRQLQQLKKLLVLQOELEFTEELQT 106  
+ E N + + Q EE E + S K ++ L + LQ+E +  
Sbjct: 1281 LTEEQNTLWISFQATHQLEKENQIKSMKA----DIESLVTEKEALQKEGNGQQAASE 1336

Query: 107 SYYSLRQYQSILEKQTSIDLVLHHHCKLKEDEVILYEEEMGNHNTGKHLHAQEQLAL 166  
+ Q + L + + + L+ K K+ E+ +++ N + L++++ A  
Sbjct: 1337 KESCITQLKELSENINAVTLMKEELKEKKVEISSLSKQLTDLNVQLQNSISLSEKEAA- 1395

Query: 167 AGDKIASLERSLNLRYDKYQSSLSNIELLECQVKMLQGELEGGIMGQEPENKGDHRSKVIY 226  
I+SL + Y ++ L ++ L +V L E + Q + S+ +  
Sbjct: 1396 ----ISSLRKQ---YDEEKCELLDQVQDLSFKVDTLSKEKISALEQVDDWSNKFSEWK-K 1447

Query: 227 TSPCMIQEHQETQKRLS-----EVMQKVSQDDLIQEL--RNK-LACSNALVLE--- 272  
+ HQ T K L E ++K Q + L +EL +NK C + +  
Sbjct: 1448 KAQSRTQHTQNTVKELQIQLELKSKEAYEKDEQINLLKEELDQONKRFDCCLKGEMEDDKS 1507

Query: 273 -REKALIKLQADFASCTAT---HRYPSSSEECEDIKKILKHLEQKDSQCLHVEEYQN 327  
EK L+ + S TA + + E + ++L+ +QKD E+ +  
Sbjct: 1508 KMEKESNLETELKSTARIMELEDHITQKTICIESLNEVLKNYNQORDI-----EHKE 1561

Query: 328 LVKDLRVELEAVSEQNRNIMKDMKLELDLHGLREETSABIERKKDKDI--TILQCLQEL 385  
LV+ L+ + + E+K N +K+ + L L A +E K K++ L + +E  
Sbjct: 1562 LVQKLQ-HFQELGEEKDNRVKEAEKILTLENQVYSMAELETKKKELEHVNLSVSKSEE 1620

Query: 386 QLEFTEQKLTLLKDKFLQEKDEMLOLEKKLTQVQNSLLKKEKELEKQOCMATELEMTV 445  
+L+ E + L+ + + E+ ++ E+K+ ++ LL + +E +Q TE ++  
Sbjct: 1621 ELKALEDR---LESES-AAKLAELKRAEQKIAATKQLLSQMEKEEQYKKGTEHSLSE 1676

Query: 446 KEAKQDKSKEAECKALQAEVQKLNKNSLEAKQOERLAAQQAACK-EAALAGCHLEDTQ 504  
K + +E E L+ +++ ++S E RA AA + EEA GC + +  
Sbjct: 1677 LNTKLQE-REREVHILEEKLSVESSQSETLIVPRSAKNVAAYTEQEEADSQGCVCQTYE 1735

Query: 505 RKLQKGLLLDKQKADTIQELQRELQMLQKSSMAEKEQTSNRKRVLESLSEALARKLE 564  
K+ +L + + + LQR Q +KE +++ + R + +E ++L A K  
Sbjct: 1736 EKIS---VLQRNLTEKEKLLQVRGO--EKEETVSSHFM--RCQYERLIKLEHAEAKQH 1788

Query: 565 NSDKERQLOKTVAEQDMKMDMLDRIKHQHQREG--SIKCK--LE---EDLQ-----E 611  
LQ+ + E+ + K + ++ +H +E G +I+ K LE +D+Q E  
Sbjct: 1789 EDQSMIGHLQEELEENKKYSLIV--AQHVEKGGKNNIQAKQNLNVFDDVQKTLQKE 1846

Query: 612 AT-KLEDKREQLKKSKEHEKLMEG-ELEALQEFKKDKTKLENSR---KLEENENL 665  
T ++LE K ++L +K + E+ L ++K + R +L EEN  
Sbjct: 1847 LTCQILEQKIKELDSCLVRQKEVHRVEMEELTSKYELQALQMDGRNKPTELLEENTEE 1906

Query: 666 RAELOCCSTQLESSLN-KYNTSQOVIQDLNKEIALQKESLMSLQALQDKALQKEKHYLOT 724  
++ +L S++ ++N + + + +E + ++ LQ L + L+KE H +  
Sbjct: 1907 KSKSHLVQPKLLSNMEAQHNDLEFKLAGAEREKQKLGKIEIVRLQKDL-RMLRKE-HQQEL 1964

Query: 725 TITKEAYDALSRKSAACQDDLTQALEKLNHVTSETKSLQOQSLTQOTQEKKAQLEEEIAYE 784  
I K+ YD R+ Q+ + LE L H ++ + ++ TQ +K+ +LE I +  
Sbjct: 1965 EILKKEYDQ-EREKIKQEQ--EDLE-LKHNSTLKQLMREFNTQLAQKEQLEMTI---K 2017

Query: 785 ERMKKLNTLRLKRGFHOSELELVHAFDRKLEMSQVLOWQKHQNDLKMALAAKEEQLR 844  
E + K +L HQE E + KK+ E + + K+++ ++L A+EE++  
Sbjct: 2018 ETINKAQEVEAEELSHQE---ETNQLLKKIAEKDDDLKRTAKRYE---EILDAREEEMT 2071

Query: 845 EFQEEMAALKENLLEDDKEPCCLPQMSVP-KDTCRLYRGNDQIMTNLEQWAKQOKVANER 903  
++ E L + ++ L Q P D + ++ T L Q K ++ K  
Sbjct: 2072 AKVRDLQTLQLEELQKYYQK--LEQEENPGNDNVTIM---ELQTQLAQ--KTLTISDSK 2123

Query: 904 LGNQ-LREQVNYIA-KLSGEKDLHLSVMV-HL 932  
L Q REQ++ + +L + +++ V HL  
Sbjct: 2124 LKEQEFREQIHNLDRLLKKEKNVYATTVGHL 2155

Score = 280 (42.0 bits), Expect = 5.2e-20, P = 5.2e-20  
Identities = 209/938 (22%), Positives = 432/938 (46%)

Query: 3 DEAGERDREVS-SLNSKLLSLQDIKN-LHDVC-KRQKTLQDNQCMEEAM-NSSHDKK 58  
++ ++ +E+ +L KLL + +K L + + +K Q N +E A NS+  
Sbjct: 957 EKVQKAKEMQETLKKLLDQEAKKKELENTALELSQKEKQFNKMLEMAQANSAGISD 1016

Query: 59 QAQALAFEESEVEFGSSKQCHLRQLQQLKKLLVLQOELEFTEELQTSYYSLRQYQSIL 118  
L + E + S + H R+L + + + ++L EELO + ++ +  
Sbjct: 1017 AVSRLETNQE-QIESLTVHRRLENDV---ISIWKKLNQQAEELO-EIHEIQLEK-- 1069

Query: 119 EKQTSDLV--LLHHHCKLKE-DEVILYEEEMGNHNTGKHLHAQEQLALAGDKIASLE 175  
E++ ++L +L C+ +E ++ I + +E G + T +L +Q + + +A E  
Sbjct: 1070 EQEVAELKQKILLFGCEKEEMNKEITWLKEEGVKQDITLNELOEQKQSAHVNSLAQDE 1129

Query: 176 RSLNLYRDKYQSSLSNIELLECQVKMLQGELEGGI--MGQEPENKGDHRSKVIYRTPCMQ 233  
L + +K + L N L E LQ +L + + +E + K ++ T+ Q

Sbjct: 1130 TKLKAHLEKLEVDL-NKSLKENT--FLQEQLVELKMLAEEDKRKVSELTSKLKTTDEEFQ 1186

Query: 234 E---HQETQKRLSEVWQKVSQDDLIQELRNKL--AC--SNALVLEREKALIKLQADFA 285  
H+++ K L + K + L +EL +L C + AL+ + LI + +

Sbjct: 1187 SLKSSHEKSNKSLD---KSLEFKKLSLELATQLDICCCKTEALLEANTELINISSKT 1243

Query: 286 SCTATH-RYPPSSSEECEDIKKILKHLQEOKDSQCLHVEEYQNLVKDLRVELEAVSEQKR 344  
+ + + + + + + + I + + + Q + E QN + + + E + K

Sbjct: 1244 NAILSRISHCQHRITTKVEKALLINTCTVSELAQLRQLTTEEQNTLNISFQQATHQLEKE 1303

Query: 345 NIMKDMMLKELD-LHGLREETSATIERKDKDITILQCRLEQLQLETTET-QKLTLLKDKF 402  
N + K M K + + L + E + + + + + + + L + E + E + + TL K + +

Sbjct: 1304 NQIKSM-KADIESLVTEKALQKEGNGQQAASEKESCITQLKKELSENINAVTLMKEE- 1361

Query: 403 LQEKDEMLOLEKLLTQVQNSLLKKEKELEKQCCMATELEMTVKEAQDKSKEAECKALQ 462  
L + EK + L K + LT + N L + L + + + + L E K + + + + L

Sbjct: 1362 LKEKKVEISSLSKQLTDL-NVQLQNSISLSEKAAISSLRKQYDEKCELLOQVQ--DLS 1418

Query: 463 AEVQKLKNSLEEAQOERLAAQAAQCKEEAALAGCHLEDTORLKLQGLLLDKQKA---- 518  
+ V L A + Q + + + + K + A + + T + L Q L L + A

Sbjct: 1419 FKVDLSKEKISALEQVDDWSNKFSEWKKKAQSRFTQHONTVKELQIQLEKSKAEYKED 1478

Query: 519 DTIQELQRELMQLKESMAEKEQTSNRKRVEELSLELSEALRKLENSDKEKRQLQKTV 578  
+ I L + EL K + E + + + E + L + L + + + L + + +

Sbjct: 1479 EQINLLKEELDQQKRFDCLEKGEEDDKSKMEKESNLET---ELKSQTARIMELEDHIT 1535

Query: 579 EQDMKMDMLDRIKHQHREQSGSIKCK-LEEDLQEAATKLEDKREQLKSKSEHEKLMGEL 637  
+ + + + + + + K + + Q I + K L + L Q + L E + K + + K + + E + E + +

Sbjct: 1536 QKTIEIESLNEVLKN-YNOQKDIEHKELVQKLQHFQELGEEKONRVKEAEKILTLENQV 1594

Query: 638 EALRQEFKKDKTLKENSRLKEEENENLRAELQCCSTQLES-SLNKYNTSQQVQIDLNKE 696  
+ + + E + K K L + + + + E L + A L + + LES S K + + + +

Sbjct: 1595 YSMKAELETKKKELEHVNLSVKSKEELKA-LE---DRLESESAALK---AELKRKAQK 1647

Query: 697 IALQKESLMSLAQOLDKALQKEHYLQTTITKEAYDALSRKSAACQDDLTQALEKLNHVT 756  
IA K + L + S Q + + + KE + Y + T + L + K + + + + EK L V

Sbjct: 1648 IAAIKKQLLS---QME---EKEDQYKKG--ESHLSLNTKLQEREREVHILEKLSKVE 1699

Query: 757 S---ET---KSLQOSTQTEKKAQLEEEII-AYEERMKKLNTLRLKRGFHOSELEV 808  
S ET + S + T + + + A + + YEE + L L L E E +

Sbjct: 1700 SSQSETLIVPRSAKNVAAYTEQEADSGCVQKTYEKBISVLQRNLT-----EKEKLL 1752

Query: 809 HAFDKLEMSQVQLWQKHQNDLKMALAEQLEFQEMAALKNLEDDKEPCCLP 868  
+ + EE + + Q + Q L L L E + E Q + L + E L E + K + +

Sbjct: 1753 QRVGQKEETVSSHFMRCQYQERLIKLEHAEAKQHEQDSMIGHLQEELEENKKYSLIV 1812

Query: 869 QMSVPKDTCLRLYRGNDQIMTNLEQ-WAKQKQVANER-LGNQLREQ-VNYIAKLSGEKDH 925  
V K + + N Q NLE + QK EK L Q + EQ + + + +

Sbjct: 1813 AQHVEKEGGK---NNIAQKQNLNLFDDVQKTLQEKELTCQILEQKIKELDSCLVRQKEV 1869

Query: 926 HSV-MVHLQENKKLK 940  
H V M L + + KL +

Sbjct: 1870 HRVEMEELTSKYELQ 1885

Score = 227 (34.1 bits), Expect = 2.5e-14, P = 2.5e-14  
Identities = 160/716 (22%), Positives = 318/716 (44%)

Query: 233 QEHQETQKRLSEVWQKVSQDDLIQE-LRNKLACSNALV-LEREKALIKL-QADFASCTA 289  
+ E + TO + + + + V + L + + L S + + L R + L + D S TA

Sbjct: 53 RESGDTQSFQAKLQLRVPSVESLFRSPIKESLFRSSSKESLVRTSSRESLNRDLDSSTA 112

Query: 290 THRYPPSSSEECEDIKKILKHLQEOKDSQCLHVEEYQNLVKDLRVELEAVSEQKRNMKD 349  
+ P E ED+ L + + + Q L + + R + + + + +

Sbjct: 113 SFDPPSDMDSEADLVGNSDSLNKEQLIQLRL--RMERSLSSYRGKYSSELVTAYQMLQRE 170

Query: 350 MMKLELDLHGLREETSATIERKDKDIT-ILQCRLEQLQLETTETQKLTLLKDKFLQEKDE 408  
KL + G + + + + DK + I + R + ELQ + + L + D L + EKD +

Sbjct: 171 KKKLQ-----GILSQS-----QDKSLRRIAEELR-EELQMDQAKHLQEEDFASLEEKDQ 219

Query: 409 MLQLEKLLTQVQNSLLKKEKELEKQCCMATELEMTVKEAQDKSKEAECKALQAE---V 465  
+ L + + + + L + + + + + LE + + + + E + + + + V

Sbjct: 220 YISVLQTVSLLKQRLRNGPMNVOLKPLP-QLEPQAEVFTKEENPESDGPVVEDGTSV 278

Query: 466 QKLKNSLEEAQOERLAAQAAQCKEEAALAGCHLEDTORLKLQGLL-LDKQKADTI 521  
+ L + + K + QE L + + Q KE + L E Q + L + L + K K +

Sbjct: 279 KTLQLEQVRKQENLLKRCKETIQSHKEQCTLLTSEKALQEQDLERLQLEKIKDLHM 338

Query: 522 QELQRELMQLKESMAEKEQTSNRKRVEELSLELSEALRKLENSDKEKRQLQKTVAEQD 581  
E + + L + + E + + + E + + E L E + R K + Q

Sbjct: 339 AEKTKLITQLDAKNLIEQLEQDKGMVIAETKQMHETLEMKEEIIAQLRSRIKQMTTQG 398

Query: 582 MKNMMDLRIKHQHREGQSIKCKLEEDLQEAT-KLEDKREQLK---KSKEHEKL-MEGE 636  
 ++ ++ ++ E+ + +EA KL + EQ+K K+ E E++ ++ E  
 Sbjct: 399 EELREQEKSERAAFEELEKALSTAQKTEEARRLKAEMDEQIKTIKTSEERISLQOE 458

Query: 637 LEALRQEFKK-KDKTLKENSRLKEEENENLRAELQCCSTQLESSLNKYNTSQQVIQDLNK 695  
 L ++QE K+ +E KL++ +E EL +L L T ++ Q+ K  
 Sbjct: 459 LSRVKQEVVDMKKSSEEQIAKLQKLHEK---ELARKEQELTKKLQ---TREREFQEQMK 512

Query: 696 EIALQKESLMSLQAQLDKALQKEKHYLTITKEAYDALSRKSAACQDDLTQALEKLN-H 754  
 +AL+K L+ +K Q+ + +K+A S DL Q E  
 Sbjct: 513 -VALEKSQSEYLIKISQEKEQESLAELELQKKAILTESENKLR---DLQQAETRYRTR 568

Query: 755 VTSETKSLQSLTQTQEKKAQLEEEIIAYEERMKKLNTLRLKRGFHOESELEV--HAFD 812  
 + SL++SL OE K Q ++ + E K N E+ + H+ +ELE H D  
 Sbjct: 569 ILELESSLEKSL---QENKNQSKDLAVHLEAEKNKHKEITVMVEKHK-TELESXHQD 624

Query: 813 KALEEMSCQVLQWQKQNDLKMALAAKEEOLRE-----FOEMAALKENLLED-DK 862  
 E QVL+ +Q+Q +++ L K EQ +E FO + + E LE D  
 Sbjct: 625 ALWTE-KLQVLK--QQYQTEMEKREKCEQEKETLLKDKEIIFQAHIEEMNEKTLEKLOV 681

Query: 863 EPCCLPQWSVPKDTCLRYRGNDQIMTNLEQWAKQKQVANERKLNQRLQEVNYIAKLSGEK 922  
 + L S+ + + + + L Q ++L ++ EQ N+ +  
 Sbjct: 682 KQTELE--SLSSELSEVLKARHKLEELSVLKDQTKMKQLEAKMDEQKNHHQQQVDSI 739

Query: 923 DHLHSMVMHLOQENKKLKEIEEKKM 948  
 H V + Q+ K LK +I + ++  
 Sbjct: 740 IKEHEVSI--ORTEKALKDQINQLEL 763

Score = 183 (27.5 bits), Expect = 1.3e-09, P = 1.3e-09  
 Identities = 132/584 (22%), Positives = 251/584 (42%)

Query: 409 MLQEEKLKTQVNSLLKKEKELEKQCCMATELEMTVEAK-QDKSKEAECKALQAEVQK 467  
 M ++L++K+++ Q L + + +T M + + + ++ E + + Q  
 Sbjct: 1 MFKKLKQKISIEEQOQLQALAPAQASSNSSTPTRMSRTSSFTQLEDGTPNRESGDTQS 60

Query: 468 LKNSLE-EAQQERLAAQQAQCKEEAALAGCHLEDTQRKLQKGLLLDKQKA--DTIQEL 524  
 L+ E L + + + + +R+ L LD A D ++  
 Sbjct: 61 FAQLQLRVPSVESLFRSPIKESLFRSSSKESLVRTSSRESLNRDLDSSTASFDPPSDM 120

Query: 525 QRELQMLQKESMAEKEQTSNKRVEELSL-----ELSEALRKLNSDKERQLOKTVAE 579  
 E + L S KEQ R R E SL + SE + + +EK++LQ +++  
 Sbjct: 121 DSEAEOLVGNSDSLNEQLIQLRLRMERSLSSYRGYSELVTAYQMLQREKKKLOGILSQ 180

Query: 580 -QDMKNMMDLRIKHQHREGQSIKCKLEE---DLQEATK---LLEDKREQLKKSKEHEKL 632  
 QD + + + + +Q+ K EE L+E + +L+ + LK+ + +  
 Sbjct: 181 SQKSLRRIAELELQMDQAKKHLQEEFDASLEEKDQYISVLQTVSLLKQRLRNGPM 240

Query: 633 MEGEALRO-EFKKKDKTLKENSRLKEE---ENENLRAELQCCSTQLESSLNKYNTSQQ 688  
 L+ L Q E + + T +EN E E+ L+ +++ N ++  
 Sbjct: 241 NVDVLKPLPQLEPAEVFTKEENPESDGEFVVEDGTSVKTLETQQRVQRQENLLKRCKE 300

Query: 689 VIQDLNKEIALQKESLMSLQAQLDKALQKEKHYLTITKEAYDALSRKSAACQDDLTQA 748  
 IQ ++ L +LQ QLD+ LQ E ++ E +++ A +L +  
 Sbjct: 301 TIQSHKEQCTLLTSEKALQEQDLERLQ-ELEKIKDLHMAEKTKLITQLRDA--KNLIEQ 357

Query: 749 LEK-LNHVTSETKSLQSLTQTQEKKAQLEEEIIAYEERMKKLNTLRLKRGFHOESELE 807  
 LE+ V +ETK + + +T E K EEEI R+K++ T+ +LR Q+ + E  
 Sbjct: 358 LEQDKGMVIAETK---RQMHTLENK---EEETIAQLRSRIKQMTTQGEELR--EQKEKSE 409

Query: 808 VHAFDKKLEEMSCQVLQWQKQNDLKMALAAKEEOLREFQ----EEMAALKENLLEDDKE 863  
 AF EE+ + QK + K+ A +EQ++ + EE +L++ L +E  
 Sbjct: 410 RAAF---EELEKALSTAQKTEEARRLKAEMDEQIKTIKTSEERISLQOELSRVKQE 465

Query: 864 PCCLPQWSVPKDTCLRYRGNDQIMTNLEQ-WAKQKQVANERKLNQRL-----EQVNYIAK 917  
 + S + +L + +++ + EQ K+ + + Q++ Q Y+ K  
 Sbjct: 466 VVDVMKKSSEEQIAKLQKLHEKELARKEQELTKKLQTREREFQEQMKVALEKSQSEYL-K 524

Query: 918 LSGEKDHLHSMVMH-LQENKKLKEIEEK---KMKAEENTRLCTKALGPSRTESTOREK 972  
 +S EK+ S+ + L+ + K+ E E K + +AE R L S +S Q K  
 Sbjct: 525 ISQEKEQESLAELELQKKAILTESENKLRDLQQAETRYRTRILELESSLEKSLQENK 584

Pedant information for DKFZphtes3\_lg13, frame 1  
 -----

Report for DKFZphtes3\_lg13.1

{LENGTH} 1007  
 {MW} 117480.77  
 {PI} 5.90

[HOMOL] TREMBL:AF092090\_1 product: "cp151"; Rattus norvegicus cp151 mRNA, partial cds.  
0.0  
[FUNCAT] 30.03 organization of cytoplasm [S. cerevisiae, YDL058w] 5e-15  
[FUNCAT] 08.07 vesicular transport (golgi network, etc.) [S. cerevisiae, YDL058w] 5e-15  
[FUNCAT] 09.10 nuclear biogenesis [S. cerevisiae, YDR356w] 1e-11  
[FUNCAT] 30.04 organization of cytoskeleton [S. cerevisiae, YDR356w] 1e-11  
[FUNCAT] 03.22 cell cycle control and mitosis [S. cerevisiae, YDR356w] 1e-11  
[FUNCAT] 30.10 nuclear organization [S. cerevisiae, YKR095w] 1e-08  
[FUNCAT] 11.04 dna repair (direct repair, base excision repair and nucleotide excision repair) [S. cerevisiae, YKR095w] 1e-08  
[FUNCAT] 99 unclassified proteins [S. cerevisiae, YLR309c] 1e-08  
[FUNCAT] 1 genome replication, transcription, recombination and repair [M. jannaschii, MJ1322] 4e-06  
[FUNCAT] 09.13 biogenesis of chromosome structure [S. cerevisiae, YLR086w] 9e-06  
[FUNCAT] 03.04 budding, cell polarity and filament formation [S. cerevisiae, YHR023w MYO1 - myosin-1 isoform] 3e-04  
[FUNCAT] 08.22 cytoskeleton-dependent transport [S. cerevisiae, YHR023w MYO1 - myosin-1 isoform] 3e-04  
[FUNCAT] 03.25 cytokinesis [S. cerevisiae, YHR023w MYO1 - myosin-1 isoform] 3e-04  
[FUNCAT] 98 classification not yet clear-cut [S. cerevisiae, YJR134c] 5e-04  
[EC] 3.6.1.32 Myosin ATPase 1e-16  
[PIRKW] nucleus 3e-10  
[PIRKW] phosphotransferase 6e-09  
[PIRKW] duplication 2e-06  
[PIRKW] citrulline 2e-12  
[PIRKW] tandem repeat 1e-16  
[PIRKW] endocytosis 2e-13  
[PIRKW] heart 8e-13  
[PIRKW] transmembrane protein 1e-13  
[PIRKW] serine/threonine-specific protein kinase 6e-09  
[PIRKW] zinc finger 2e-13  
[PIRKW] metal binding 2e-13  
[PIRKW] DNA binding 4e-12  
[PIRKW] muscle contraction 1e-16  
[PIRKW] acetylated amino end 1e-11  
[PIRKW] actin binding 1e-16  
[PIRKW] mitosis 5e-15  
[PIRKW] microtubule binding 5e-15  
[PIRKW] ATP 1e-16  
[PIRKW] thick filament 1e-16  
[PIRKW] phosphoprotein 4e-16  
[PIRKW] skeletal muscle 2e-14  
[PIRKW] calcium binding 2e-12  
[PIRKW] alternative splicing 1e-16  
[PIRKW] coiled coil 1e-16  
[PIRKW] P-loop 1e-16  
[PIRKW] heptad repeat 3e-10  
[PIRKW] methylated amino acid 1e-16  
[PIRKW] immunoglobulin receptor 2e-06  
[PIRKW] peripheral membrane protein 2e-13  
[PIRKW] cardiac muscle 8e-13  
[PIRKW] hydrolase 1e-16  
[PIRKW] microtubule 3e-10  
[PIRKW] muscle 8e-13  
[PIRKW] EF hand 2e-12  
[PIRKW] cytoskeleton 2e-15  
[PIRKW] hair 2e-12  
[PIRKW] calmodulin binding 2e-13  
[PIRKW] Golgi apparatus 3e-10  
[SUPFAM] myosin heavy chain 1e-16  
[SUPFAM] conserved hypothetical P115 protein 1e-07  
[SUPFAM] centromere protein E 5e-15  
[SUPFAM] unassigned Ser/Thr or Tyr-specific protein kinases 6e-09  
[SUPFAM] calmodulin repeat homology 2e-12  
[SUPFAM] myosin motor domain homology 1e-16  
[SUPFAM] alpha-actinin actin-binding domain homology 2e-07  
[SUPFAM] plectin 2e-07  
[SUPFAM] trichohyalin 2e-12  
[SUPFAM] pleckstrin repeat homology 8e-08  
[SUPFAM] ribosomal protein S10 homology 2e-07  
[SUPFAM] giantin 3e-13  
[SUPFAM] protein kinase homology 6e-09  
[SUPFAM] protein kinase C zinc-binding repeat homology 8e-08  
[SUPFAM] kinesin motor domain homology 5e-15  
[SUPFAM] human early endosome antigen 1 2e-13  
[SUPFAM] M5 protein 1e-07  
[PROSITE] LEUCINE\_ZIPPER 7  
[PROSITE] MYRISTYL 2  
[PROSITE] CAMP\_PHOSPHO\_SITE 2  
[PROSITE] CK2\_PHOSPHO\_SITE 20

[illegible]

```
COILS .....
SEQ  NEKLGNQLREQVNYIAKLSGEKDHLSVMVHLQENKKLKEIEKKMKAENTRLCTKAL
SEG  .....xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx
PRD  hhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh
COILS .....CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC

SEQ  GPSRTESTQREKVCGLGWKGLPQDMGQRMDLTKYIGMPHCPGSSSYC
SEG  .....
PRD  cchhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh
COILS .....

```

## Prosites for DKFZphtes3\_lg13.1

PS00001	52->56	ASN_GLYCOSYLATION	PDOC00001
PS00001	684->688	ASN_GLYCOSYLATION	PDOC00001
PS00004	240->244	CAMP_PHOSPHO_SITE	PDOC00004
PS00004	415->419	CAMP_PHOSPHO_SITE	PDOC00004
PS00005	74->77	PKC_PHOSPHO_SITE	PDOC00005
PS00005	110->113	PKC_PHOSPHO_SITE	PDOC00005
PS00005	238->241	PKC_PHOSPHO_SITE	PDOC00005
PS00005	290->293	PKC_PHOSPHO_SITE	PDOC00005
PS00005	392->395	PKC_PHOSPHO_SITE	PDOC00005
PS00005	396->399	PKC_PHOSPHO_SITE	PDOC00005
PS00005	444->447	PKC_PHOSPHO_SITE	PDOC00005
PS00005	503->506	PKC_PHOSPHO_SITE	PDOC00005
PS00005	544->547	PKC_PHOSPHO_SITE	PDOC00005
PS00005	566->569	PKC_PHOSPHO_SITE	PDOC00005
PS00005	600->603	PKC_PHOSPHO_SITE	PDOC00005
PS00005	650->653	PKC_PHOSPHO_SITE	PDOC00005
PS00005	655->658	PKC_PHOSPHO_SITE	PDOC00005
PS00005	735->738	PKC_PHOSPHO_SITE	PDOC00005
PS00005	876->879	PKC_PHOSPHO_SITE	PDOC00005
PS00005	968->971	PKC_PHOSPHO_SITE	PDOC00005
PS00006	39->43	CK2_PHOSPHO_SITE	PDOC00006
PS00006	53->57	CK2_PHOSPHO_SITE	PDOC00006
PS00006	68->72	CK2_PHOSPHO_SITE	PDOC00006
PS00006	116->120	CK2_PHOSPHO_SITE	PDOC00006
PS00006	190->194	CK2_PHOSPHO_SITE	PDOC00006
PS00006	250->254	CK2_PHOSPHO_SITE	PDOC00006
PS00006	296->300	CK2_PHOSPHO_SITE	PDOC00006
PS00006	439->443	CK2_PHOSPHO_SITE	PDOC00006
PS00006	444->448	CK2_PHOSPHO_SITE	PDOC00006
PS00006	471->475	CK2_PHOSPHO_SITE	PDOC00006
PS00006	520->524	CK2_PHOSPHO_SITE	PDOC00006
PS00006	536->540	CK2_PHOSPHO_SITE	PDOC00006
PS00006	566->570	CK2_PHOSPHO_SITE	PDOC00006
PS00006	576->580	CK2_PHOSPHO_SITE	PDOC00006
PS00006	650->654	CK2_PHOSPHO_SITE	PDOC00006
PS00006	674->678	CK2_PHOSPHO_SITE	PDOC00006
PS00006	804->808	CK2_PHOSPHO_SITE	PDOC00006
PS00006	888->892	CK2_PHOSPHO_SITE	PDOC00006
PS00006	963->967	CK2_PHOSPHO_SITE	PDOC00006
PS00006	968->972	CK2_PHOSPHO_SITE	PDOC00006
PS00007	135->143	TYR_PHOSPHO_SITE	PDOC00007
PS00008	207->213	MYRISTYL	PDOC00008
PS00008	599->605	MYRISTYL	PDOC00008
PS00029	83->105	LEUCINE_ZIPPER	PDOC00029
PS00029	90->112	LEUCINE_ZIPPER	PDOC00029
PS00029	97->119	LEUCINE_ZIPPER	PDOC00029
PS00029	104->126	LEUCINE_ZIPPER	PDOC00029
PS00029	403->425	LEUCINE_ZIPPER	PDOC00029
PS00029	410->432	LEUCINE_ZIPPER	PDOC00029
PS00029	918->940	LEUCINE_ZIPPER	PDOC00029

(No Pfam data available for DKFZphtes3\_lg13.1)

DKFZphtes3\_lk11

-----

group: cell structure and motility

DKF2phtes3\_lkl1 encodes a novel 589 amino acid protein with strong similarity to *Mus musculus* actin-binding protein (ENC-1).

Ectoderm-neural cortex-1 protein (ENC-1) is an early and highly specific marker of neural induction in vertebrates. The protein is related to the kelch family proteins and is expressed during early gastrulation in the prospective neuroectodermal region of the epiblast and later in development throughout the nervous system (NS). ENC-1 functions as an actin-binding protein organising the actin cytoskeleton during neural differentiation and development of the NS. The novel protein is highly similar to ENC-1.

The new protein can find application in modulation of cyto skeleton organisation in human testicular cells.

strong similarity to mouse ENC-1

complete cDNA, compete cds, EST hits

Sequenced by DKF2

Locus: unknown

Insert length: 3525 bp

Poly A stretch at pos. 3515, polyadenylation signal at pos. 3499

```
1 GGTGGAGAGC CGGCCGACGG GAGCCGCGGC GGAGCCTGTT GAGCTCGCGC
51 GGGCTGCCGG GAGTGGTCTC TGAGGCGGCG GCGGCGCGGG GGATCGTCTC
101 CGGCACTGGC GCACCATGTC GGTCACTGTC CATGAGACCC GCAAGTCGGC
151 GAGCAGCAGC GGGTCCATGA ACGTCACCCT CTTCCACAAG GCCTCCACCC
201 CGGACTGTGT GCTGGCCACG CTCACACCGC TTCCCAAGCA CTGCATGTTT
251 ACCGACGTCA CACTCTAGGC GGGCGACCGT GCCTTCCCTC GTCAACGTGC
301 CGTGTGGGCC GCCTCTAGCC GCTATTTTGA GGCCATGTTT AGCCATGGCC
351 TTCCGGGAGG CCGGGATGAC ACTGTCAACT TCCAGGACAA CCTGCACCCG
401 GAGGTGCTGG AGCTGCTGCT GGACTTTGCC TACTCTCAC GCATCGCCAT
451 CAACGAGGAG AACGCTGAGT CACTGTGGA GGCAGGCGAC ATGCTGCAGT
501 TCCACGATGT GCGGGATGCT GCCGCCGAGT TCCTGGAGAA GAACCTTTTC
551 CCCTCCAACCT GCCTGGGCGT GATGCTGCTC TCGGAGCGCC ACCAGTGCCG
601 CCGGCTGTAT GAGTTCTCCT GCGCATGTGT CTTGTGCCA TTTGAGACGG
651 TGAGGACAGG CGAGGACTTC AACAGCTGTG CCAAGCACAC ACTCTGGAC
701 CTCATCTCGA GTGATGAGCT GGAGACCGAG GACGAGCGGG TGGTCTCGA
751 GGCCATCTCT CAGTGGGTGA AGCAGCACT GGAGCCACGG AAGGTCCACT
801 TGCCCGAGCT CCTCCGACG GTGCGTCTGG CTTGTGTCGC GTCCGACTGC
851 CTGAGGAGG CCGTCTCCAG CGAGGCCCTC CTCATGGCAG ACGAGCGCAC
901 CAAGCTTATC ATGGATGAGG CCCTGCGCTG CAAGACGAGG ATCCTGCACA
951 ATGATGGCGT GGTCAACGAG CCCTGTGCCC GGCCACGCAA GGCGGGCCAC
1001 ACGCTACTCA TCCTGGGGGG CCAGACCTTC ATGTGTGACA AGATCTACCA
1051 GGTGGACAC AAGGCCAAGG AGATCATCCC CAAGGCCGAG CTGCCACGCC
1101 CCCGGAAGGA GTTCAGCGCC TCAGCGATCG GCTGCAAGGT CTATGTGACG
1151 GGGGGCAGGG GCTCCGAGAA CGGGGTCTCC AAGGATGTCT GGTGTACGAC
1201 CACCGTACAT GAGGAATGGT CCAAGGCGGC GCCATGCTGT ATTGCCCGCT
1251 TTGGCCATGG CTCAGCTGAG CTGGAGAACT GCCTCTATGT GGTGGGGGGA
1301 CACACATCCC TGGCAGGGGT CTTCCCGGCC TCGCCTTCTG TCTCCCTGAA
1351 ACAAGTGGAG AAATACGACC CTGGGGCCAA CAAGTGGATG ATGGTGGCCC
1401 CCTTGCAGGA TGGCGTCAGC AATGCCGAGG TGGTGAATGC CAAGCTGAAG
1451 CTCCTTTGTT TCGGAGGAAC CAGCATCCAC CGGGACATGG TGTCCAAGGT
1501 CCAGTGCTAT GACCCCTCGG AGAACAGGTT GACGATCAAG GCCGAGTGCC
1551 CCCAGCCTTG GCGGTACACA GCGCTGCGCG TCCTGGGCAG CCAGATCTTC
1601 ATCATGGGAG GTGACACGGA ATTACAGGCC GCCTCGGCCCT ACCGCTTTGA
1651 CTGTGAGACC AACCAGTGGG CGCGGATTGG GGACATGACT GCCAAGCGCA
1701 TGTCTTGCCA TGCCCTGGCT TCCGGCAACA AGCTCTATGT GGTCCGGGGC
1751 TACTTTGGGA CCCAGAGGTG TAAGACTCTG GACTGTATGT ACCCCACTTC
1801 AGATACATGG AACTGCATCA CCACAGTCCC CTACTCACTT ATCCCCACGG
1851 CCTTTGTGAG CACTTGGGAG CACTTGCCTG CGTGAGGAGC ACCTGTCTGAG
1901 CCCAGCCAGA CCGCGGCTTT CAGTGTCAAC GCCTGGCCTT GCTTGTCTGC
1951 CACAGCGGGA GCTAAGCCGG CCCTGGGCCA GCACTCCGAG AGTGGGAAGG
2001 GGCCCTGCCA GCTCTGGGGA GCAGACGCTT TGGGCTGTTT TGAGCTTTAG
2051 GCAAGAGAAG AGAAGCATCT CTTGCATCCG TGCCCTGGGG GGCCTCTTCA
2101 GCTTTGCACT GGTTTGTGGG AAGACATACC TCCCAGAGGG GCATGGACTG
2151 CCACCAAGAG TGACCTTGGC GTCCGGGAGA AGGACACTTG CAGAGCCTTG
2201 AGATCACCTG TTTGGCAGGT CCTGGACTGG GGCCGGGCGG GCAGGGGCGG
2251 GGAGGCGCCC CGGGTGGGCT TTTGGGCTCT TGCCATCTAT AGATGGGTGC
2301 CCCTCTTGGC CTGCCCTGCT GGGGCTCTAC TGCCATCTAT AGATGGGTGC
2351 CTGGGCTGGG GAAACTAGGT TCCCAGGGGT TGAGACGAGA AAGGTGACCA
2401 AGACAGATTT TTTAAGGTGC AGAACTGCA GGGGGGCTC AGTGACATCC
2451 ATGAGGCCCTT ATTAAGCAAAG GACACCCAGA CCTCCAAGGT TTGTGGGCCC
2501 CTTCCACAAA GCTGTAAGTC CCAGCCCACT TACTCAGGGC CTTGCTCAGT
2551 GCTGTGGCCC GGTGGGGACA CAGTTGCTCG TGGCCACTCA GTGGAGCTGG
2601 GCCTGCAGCA GACTCAAGGC TCCGAGTGCC CTGGGGGTCA CCCCCTCCCT
2651 CCCCCTCTCA GAGCCACCC TGAGAGGCGC CAGTGACCCC CATGGCACAC
2701 ACCTGCCAAC AGCACTGGGG GCTTCTCCCC AGGAGACCAC GCTGCCCTCC
```



```

2751 AAGACCAGGA GCAGCTGTGA GCTGGAGACA GCAGAGGGAC CCCAGGGTGT
2801 CCCCTGCAGA TCCCACCAGG GCCGCATCCA TCTCAGTGTG GAGGACAGTG
2851 ACGGGACCCCT CACCATCCTC TTGCGTTTTG GCCCCCATTT GCTCCCTGAG
2901 CTCCAAGATA AGAATGGCCC CGAGAGAACT GCTGAACATT TGTTCATTGC
2951 TGTCACCTCC TGAGTCACTG GGGTCCCTCA CCAGCACCTC CCTGACACCT
3001 GGGCTATGGA GAGGTGGCG CCTGTCACTG ACCATCCTAA TGCCTCTCGC
3051 TCACTCCCAA GCCACCATTT GAGAGGGAGG GGTGTTGGTG CCCTGACAGG
3101 GACTGGGCAG GGTGTCCAAA CTTGGGGCTT CCCAGGCACC TGCAGTGTGA
3151 ACACTGCTTG GCTGGCTCAA GATTAGGGCC CGGAGGGGGG CTGTGCACAT
3201 ACCAGTTACT TAAGCAGCCA CGAGTGTCCC CCATGCCTTG GTGCGGGTCC
3251 TGGAGGCCTC TTGGGGGTGG GACCTTTGGG CAGGGTTTGC CCACTGACGC
3301 GCCCGCCATG GGGCACTGGC TGCATGGGGC TCCTTGGACC CTGTAGACCC
3351 AGCAGGAGCC TGGCCGCGGG GACTGCAGGG AGGGTGCCTG GACCCGTGGG
3401 GTTGCTTCAT TGAGATAAAG CACACTTATC ACATAGCACA AAGGACGTGC
3451 CATGGTGCTT TCCCCAAAAG TTGTGTGTCT TTTATCAGTT TTCTAACTTA
3501 ATAAAAAGAG TTGAGAAAAA AAAAA

```

## BLAST Results

No BLAST result

## Medline entries

98350113:  
Cloning of human ENC-1 and evaluation of its expression  
and regulation in nervous system tumors.

97252647:  
ENC-1: a novel mammalian kelch-related gene specifically expressed in  
the nervous system  
encodes an actin-binding protein.

98234394:  
NRP/B, a novel nuclear matrix protein, associates with  
p110(RB) and is involved in neuronal differentiati

## Peptide information for frame 2

ORF from 116 bp to 1882 bp; peptide length: 589  
Category: strong similarity to known protein  
Classification: Cell structure/motility

```

1 MSVSVHETRK SRSSTGSMNV TLFHKASHPD CVLAHLNLTLR KCMFTDVTIL
51 WAGDRAFFCH RAVLAASSRY FEAMFSLGLR ESRDDTVNFQ DNLHPEVLEL
101 LLDFAVSSRI AINEENAESL LEAGDMLQFH DVRDAAEFL EKNLFPSNCL
151 GMMLLSDAHQ CRRLYEFSWR MCLVHFETVR QSEDFNSLSK DTLDLIISD
201 ELETEDERVV FEAILQWVKH DLEPRKVHLP ELLRSVRLAL LPSDCLQEA
251 SSEALLMADE RTKLIMDEAL RCKTRILQND GVVTSPCARP RKAGHTLLIL
301 GGOTFMCDKI YQVDHKAKEI IPKADLPSPR KEFSASAIGC KVVYVTGGRGS
351 ENGVSKDVMV YDTVHEEWSK AAPMLIARFG HGSAELENCI YVVGHTSLA
401 GVFPASPSVS LKQVEKYDPG ANKWMNVAPL RDGVSNAAVV SAKLKLFPVG
451 GTSIHRDMVS KVQCYDPSEN RWTIKAECQ PWRYTAAAVL GSQIFIMGGD
501 TEFTAASAYR FDCETNQWTR IGDMTAKRMS CHALASGNKL YVVGGYFGTG
551 RCKTLDYDYP TSDTWNCITT VPYSLIPTAF VSTWKHLPA

```

## BLASTP hits

Entry MMU65079\_1 from database TREMBL:  
gene: "ENC-1"; product: "actin-binding protein"; Mus musculus  
actin-binding protein (ENC-1) mRNA, complete cds.  
Score = 2402, P = 1.9e-249, identities = 440/589, positives = 513/589

Entry AF059611\_1 from database TREMBLNEW:  
gene: "NRPB"; product: "nuclear matrix protein NRP/B"; Homo sapiens  
nuclear matrix protein NRP/B (NRPB) mRNA, complete cds.  
Score = 2400, P = 3.0e-249, identities = 440/589, positives = 512/589

Entry AF010314\_1 from database TREMBL:  
gene: "PIG10"; product: "Pig10"; Homo sapiens Pig10 (PIG10) mRNA,  
complete cds.  
Score = 1745, P = 7.8e-180, identities = 335/507, positives = 403/507

Entry KELC\_DROME from database SWISSPROT:  
RING CANAL PROTEIN (KELCH PROTEIN). >TREMBL:DMRCPA\_1 product: "ring canal protein"; Drosophila melanogaster ring canal protein and ORF2 mRNA, complete cds.  
Score = 672, P = 3.9e-66, identities = 168/536, positives = 257/536

Alert BLASTP hits for DKFZphtes3\_1kl1, frame 2

No Alert BLASTP hits found

Pedant information for DKFZphtes3\_1kl1, frame 2

Report for DKFZphtes3\_1kl1.2

[LENGTH] 589  
[MW] 65923.45  
[pI] 6.10  
[HOMOL] TREMBL:MMU65079\_1 gene: "ENC-1"; product: "actin-binding protein"; Mus musculus  
actin-binding protein (ENC-1) mRNA, complete cds. 0.0  
[FUNCAT] 10.05.99 other pheromone response activities [S. cerevisiae, YHR158c]  
2e-09  
[BLOCKS] BL01016D Glycoprotease family proteins  
[PIRKW] zinc finger 1e-08  
[PIRKW] DNA binding 1e-08  
[PIRKW] transcription factor 1e-08  
[SUPFAM] POZ domain homology 3e-68  
[SUPFAM] vaccinia virus 59K HindIII-C protein 1e-15  
[SUPFAM] A55R protein 5e-29  
[SUPFAM] hypothetical protein YHR158c 4e-08  
[SUPFAM] A55R protein middle region homology 5e-29  
[SUPFAM] myxoma virus M9-R protein 1e-14  
[SUPFAM] A55R protein carboxyl-terminal homology 5e-29  
[KW] Alpha\_Beta

SEQ MSVSVHETKRSSTGSMNVTLFHKASHPCVLAHLNLTLRKHCWFTDVTLMAGDRAFPCH  
PRD ccc

SEQ RAVLAASSRYFEAMFSGHLSRSDDTVNFQDNLHPEVLELLDFAYSSRIANEENAESE  
PRD hcc

SEQ LEAGDMLQFHDVDAAAEFLEKNLFPSNCLGMMLLSDAHQCRRLYEFSWRMCLVHFETVR  
PRD hhh

SEQ QSEDFNSLSKDTLLDLISSDELETEDERVVFEAILQWVKHDLPRKVHLPPELLRSVRLAL  
PRD hhh

SEQ LPDCLQEAVSSEALLMADERTKLIMDEALRCKTRILQNDGVVTPCARPRKAGHTLLIL  
PRD cchhh

SEQ GGQTFMCDKIYQVDHKAKEIIPKADLPSPRKEFSASAIGCKVYVTGGRGSENGVSKDVWV  
PRD ccc

SEQ YDTVHEEWSKAAPMLIARFGHGSAELENCLYVVGHTSLAGVFPASPSVSLKQVEKYDPG  
PRD ccc

SEQ ANKWMVAPLRDGVSNAAVVS AKLKL FVGTSIHRDMVSKVQCYDPSENRTWIKAECPQ  
PRD ccc

SEQ PWRYTAAAVLGSQIFIMGDTPTAASAYRFDCTNQWTRIGDMTAKRMSCHALASGNKL  
PRD ccc

SEQ YVVGGVFGTQRCKTLDCYDPTSDTWNCITVFPYSLIPTAFVSTWKHLPA  
PRD ecc

(No Prosite data available for DKFZphtes3\_1kl1.2)

(No Pfam data available for DKFZphtes3\_1kl1.2)

DKFZphtes3\_in3

group: signal transduction

DKFZphtes3\_in3 encodes a novel 1196 amino acid protein with similarity to *S. pombe* Tup1 protein.

The protein contains 1 WD-40 repeat, which is typical for the beta-transducin subunit of G-proteins. The beta subunits seem to be required for the replacement of GDP by GTP as well as for membrane anchoring and receptor recognition. In addition, a RGD site is present.

The new protein can find application in modulating/blocking G-protein-dependent pathways.

similarity to Tup1p

complete cDNA, complete cds, EST hits

Sequenced by DKFZ

Locus: /map="6q24"

Insert length: 5277 bp

Poly A stretch at pos. 5267, polyadenylation signal at pos. 5244

```
1 GCTGCATAAA GCTGAGAGAT GCCTACAGCT GAGAGTGAAG CAAAAGTAAA
51 AACCAAAGTT CGCTTTGAAA AATTGCTTAA GACCCACAGT GATCTAATGC
101 GTGAAAAGAA AAAACTGAAG AAAAAACTTG TCAGGTCTGA AGAAAAACATC
151 TCACCTGACA CTATTAGAAG CAATCTTCAC TATATGAAAG AAACACAAAG
201 TGATGATCCC GACACTATTA GAAGCAATCT TCCCATATT AAAGAACTA
251 CAAGTGATGA TGAAGTGCT GCTAACACTA ACAACCTGAA GAAGAGCAGC
301 AGAGTCACTA AAAACAAATT GAGGAACACA CAGTTAGCAA CTGAAAAATCC
351 TAATGGTGAT GCTAGTGTAG AGGAAGACAA ACAAGGAAAG CCAATAAAAA
401 AGGTGATAAA GACGGTGCCC CAGTTGACTA CACAAGACCT GAAACCGGAA
451 ACTCCTGAGA ATAAGGTTGA TTCTACACAC CAGAAAACAC ATACAAAGCC
501 ACAGCCAGGC GTTGATCATC AGAAAAGTGA GAAGGCAAT GAGGGAAGAG
551 AAGAGACTGA TTTAGAAGAG GATGAAGAA TGATGCAAGC ATATCAAGTC
601 CATGTAACCTG AAGAATGCG AAGGAGATT AAGAGGAAA TAAGAAAGAA
651 ACTGAAAGAA CAGTTGACTT ACTTCCCTC AGTACTTTA TTCCATGATG
701 ACAAACTAAG CAGTGAAAAA AGGAAAAGA AAAAGGAAGT TCCAGTCTTC
751 TCTAAGCTG AAACAAGTAC ATTGACCATC TCTGGTGACA CAGTGAAGG
801 TGAACAAAAG AAGAATCTT CAGTTAGATC AGTTCTTCA GATTCTCATC
851 AAGATGATGA AATAAGTCA ATGGAACAAA GCACAGAAGA CAGCATGCAG
901 GATGATACAA AACCTAAACC AAAAAAACA AAAAAGAAGA CTAAGCAGT
951 TGCAGATAAT AATGAAGATG TTGATGGTGA TGGTGTTCAT GAAATAACAA
1001 GCCGAGATAG CCCGGTTTAT CCCAAATGTT TGCTTGATGA TGACCTTGTC
1051 TTGGGAGTTT ACATTACCG AACTGATAGA CTTAAGTCAG ATTTTATGAT
1101 TTCTCACCCA ATGGTAAAAA TTCATGGTGT TGATGAGCAT ACTGGTCAAT
1151 ATGTCAAGAA AGATGATAGT GGACGGCCTG TTTCTCTTA CTATGAAAAA
1201 GAGAATGTGG ATTATATTCT TCCTATTATG ACCCAGCCAT ATGATTTTAA
1251 ACAGTTAAAA TCAAGACTTC CAGAGTGGGA AGAACAAATT GTATTTAATG
1301 AAAATTTTCC CTATTGTGCT CGAGGCTCTG ATGAGAGTCC TAAAGTCATC
1351 CTGTTCTTTG AGATTCTTGA TTTCTTAAGC GTGGATGAAA TTAAGAATAA
1401 TTCTGAGGTT CAAAACCAAG AATGTGGCTT TCGGAAAATT GCCTGGGCAT
1451 TTCTTAAGCT TCTGGGAGCC AATGGAATG CAAACATCAA CTCAAAACTT
1501 CGCTTGACGC TATATTACCC ACCTACTAAG CCTCGATCCC CATTAAAGTG
1551 TGTTGAGGCA TTTGAATGGT GGTCAAAATG TCCAAGAAAT CATTACCCAT
1601 CAACACTGTA CGTAACTGTA AGAGGACTGA AAGTTCAGA CTGTATAAAG
1651 CCATCTTACC GCTCTATGAT GGCTCTTCAG GAGGAAAAAG GTAAACCAAT
1701 GCATTGTGAA CGTCACCATG AGTCAAGCTC AGTAGACACA GAACCTGGAT
1751 TAGAAGAGTC AAAGGAAGTA ATAAAGTGGA AACGACTCCC TGGCAGGCT
1801 TGCCGTATCC CAAACAAACA CCTCTTCTCA CTAATGCGAG GAGAACGAGG
1851 ATGTTTGTGT CTGATTTCTT CCCACAATGG AAGAATATTA GCACAGCTT
1901 GTGCCAGCCG GGATGGATAT CCAATTATTT TATATGAAAT TCCTTCTGGA
1951 CGTTTCATGA GAGAATTGTG TGGCCACCTC AATATCATTT ATGATCTTTC
2001 CTGGTCAAAA GATGATCACT ACATCCTTAC TTCATCATCT GATGGCACTG
2051 CCAGGATATG GAAAAATGAA ATAAACAATA CAAATACTTT CAGAGTTTAA
2101 CCTCATCCTT CTTTGTGTTA CACGGCTAAA TTCCATCCAG CTGTAAGAGA
2151 GCTAGTAGTT ACAGGATGCT ATGATTCCAT GATACGGATA TGGAAAGTTG
2201 AGATGAGAGA AGATTCTGCC ATATTGGTCC GACAGTTTGA TGTTCACAAA
2251 AGTTTTATCA ACTCACTTTG TTTGATACT GAAGGTATC ATATGTATTC
2301 AGGAGATTGT ACAGGGGTGA TTGTTGTTG GAATACCTAT GTCAAGATTA
2351 ATGATTGGA ACATTCAGTG CACCACTGGA CTATAAATAA GGAATTTAAA
2401 GAAACTAGAT TTAAGGGAAT TCCAATAAGT TATTGGAGA TTCAATCCAA
2451 TGGAAACGCT TTGTTAATCC ATACCAAGA CAGTACTTTG AGAATTATGG
2501 ATCTCCGGAT ATTAGTAGCA AGGAAGTTG TAGGAGCAGC AAATTATCGG
2551 GAGAAGATTC ATAGTACTTT GACTCCATGT GGGACTTTTC TGTGCTGCTG
2601 AAGTGAGGAT GGTATAGTGT ATGTTTGGAA CCCAGAAAAC GGAGAACAAG
```

2651 TAGCCATGTA TTCTGACTTG CCATTCAAGT CACCCATTGG AGACATTCTT  
2701 TATCATCCAT TTGAAAATAT GGTTCATTTC TGTCGATTGG GGCAAAATGA  
2751 GCCAATTCCT CTGTATATTT ACGATTTCCTA TGTCGCCAG CAGGAGGCTG  
2801 AATGTTTCAA ACGCTACAAT GGAACATTTC CATTACCTGG AATACACCAA  
2851 AGTCAAGATG CCCTATGTAC CTCTCCAAA CTACCCCATC AAGGCTCTTT  
2901 TCAGATTGAT GAATTTGTCC ACCTGAAAG TTCTCAACG AAGATGCAGC  
2951 TAGTAAACA GAGGCTTGAA ACTGTACAG AGGTGATACG TTCTGTGGCT  
3001 GCAAAAGTCA ACAAATCT CTCAATTACT TCACCACAG CAGTTTCCTC  
3051 ACAACAGTCT AAGTTAAAG AGTCAAACAT GCTGACCGCT CAAGAGATTC  
3101 TACATCAGTT TGGTTTCACT CAGACCGGGA TTATCAGCAT AGAAGAGAAAG  
3151 CCTTGTAAAC ATCAGGTAGA TACAGACCA ACGGTAGTGG CTCTTTATGA  
3201 CTACACAGCG AATCGATCAG ATGAACAAAC CATCCATCGG GGAGACATTA  
3251 TCCGATGTTT TTTCAAGAT AATGAAGACT GGTGGTATGG CAGCATAGGA  
3301 AAGGGACAGG AAGGTTATTT TCCAGCTAAT CATGTGGCTA GTGAACACT  
3351 GTATCAAGAA CTGCTCCTG AGATAAAGGA GCGATCCCTT CCTTTAAGCC  
3401 CTGAGGAAAA AACTAAAATA GAAAAATCTC CAGCTCCTCA AAAGCAATCA  
3451 ATCAATAAGA ACAAGTCCCA GGACTTCAGA CTAGGCTCAG AATCTATGAC  
3501 ACATTCGAA ATGAGAAAG AACAGAGCCA TGAGGACCAA GGACACATAA  
3551 TGGATACAGG GATGAGGAG AACAGCAAG CAGGCGAGAA AGTCACTCTA  
3601 ATAGAGTAAA GAATTGAAAG AAAGTTAAGA GCTGCCGAAA TGACAGAGG  
3651 TGAATATGAC AAACCAATG GAATTTCTCT TCAGAGTTCA GAATTTTCAG  
3701 ATACTAAGGA GGAAGAAAG ATCCACTACT TCTGTCTCTT ATCAATGACT  
3751 CTAGAAAAAT CAGAATCAAG TTGTGGTGGG AAAAATCAAC GTGCGCTTTG  
3801 AGTTCAGTTG TTATAACCA TTGTGACTAT TGTGGTCAA GTTATGGTA  
3851 CTTATATCTG TACTAATTGC ATCATAATTA CATTACAGT GTTGAAGAAC  
3901 TAATGAAGAA AACACTGTAA TTGCTACTCA GCAATATGTA ATAAAGGTG  
3951 TTTGCTTAT TAGGATGCT GTTAAGTAAT CATTAAATAT TATTATATTG  
4001 GTAATGGTTG TATGTGTATG GCTATGCCCA GAATATGAAG TATCTGTTTT  
4051 TGAATTCAC TTTATTTAAA AGATAAGCAG CTGACTGGGC ACGGTGCCCTC  
4101 ATGCTGTAA TCCTAGCACC TTGGGAGGCT GAGGCGAGTG GATCACCTAA  
4151 GGTGAGGAGT TCAACAACAC CAGCCTGACC AACATGGTGA AACCCCATCT  
4201 CTACTAAAAA TACAATAATC AGCCGGGTCT CTGCGCAGG ACCTGTATATC  
4251 CCATCTACTG AGGCAGGAGA ATTGCTTGAC CCAGGAGGCA GAGGTTCGAG  
4301 TGAGCCAAAG TCACGCCATT GCACTCCAGC CTGGGGGACA GAGCAAGACT  
4351 CTATCTCCAA AAAACAAAAA AGATAAGCAG CTTTAGAATA TGGCGCATTC  
4401 AAAACAGTCT CAGTAAACAA GACATTAATA GAAACAAAT TACTTTCTAA  
4451 TTAATAATTT GTGTTCTTAA AGATCAAAAT ATATAGGTAA CTTATAGAC  
4501 CTAAATATAA AGTGATTTT GGCTGGAATG GCAACAATGT TCCCAATGTC  
4551 TTTACTTTTT AAAAAGGCT TTTCATATT AAGCACATAC CTATTTTGA  
4601 GACTTACATT GTTTAATATT TATTTTAATC TTAATATTTT TACATATTA  
4651 TATTGCATTA TTTATTTTT CTAGTTCCTA GAATAATAGT GTCAATATTA  
4701 TAGACTATAT GTTTTGAAGT TGTATATTAT AATGGGATAT TCATTTTGTG  
4751 TCTTTTCTT GACTCCTTTC TCAAGTGTGT GATAAGGTCT GCTGATATAA  
4801 TATTTAACC CAGAAAGTG AAAACTAATA TAAATTAGA AAGACCTATC  
4851 CAAATAGAC AGTCAATCC ATTAATAATA GAAGTGAGAA AAACAATGTT  
4901 GGGCATTGAG GTGTAAATTT TGCCAGATG TATACCCAGT GTGAAATATC  
4951 TTCTAATAAA AATATATTTG GCTCTTATCC CTGCACATGT AGAGGCATAA  
5001 AAATTGGTAA ACATGTCCCG CTGTGTAGAA CTTTAAAAA AAGGCATTTT  
5051 TGAATAGTGT GAGTGCCACT GATAACTGTT GAAGCCTACA GCCATCCGCC  
5101 CAAAAGTCTG TTCTGATGCG ACTGAGTTTT CATTGTTCTG GATGTATAAG  
5151 TCTGTGTGTC AGGTACAGCT GGGCCAGGCC AGCTTGAGTC ACTCTGTATC  
5201 AAGCTTGTTT TTTCTGTCT TGTGAATGCA CTTGATAATT TAAAAATAAA  
5251 AATATCTGTT TCTCTGCAA AAAAAA

## BLAST Results

Entry HS32B1 from database EMBL:  
Human DNA sequence \*\*\* SEQUENCING IN PROGRESS \*\*\* from clone 32B1  
Score = 4445, E = 0.0e+00, identities = 889/889

Entry U93816 from database EMBL:  
Human exon-trapped sequence from 6q24.  
Score = 965, E = 4.0e-35, identities = 193/193

## Medline entries

No Medline entry

## Peptide information for frame 1

ORF from 19 bp to 3606 bp; peptide length: 1196  
Category: similarity to known protein

```

1 MPTAESEAKV KTKVRFEKLL KTHSOLMREK KKLKKLVRS EENISPDITR
51 SNLHYMKETT SDDPDTIRSN LPHIKETTSO DVSAANTNLL KKSTRVTKNK
101 LRNTQLATEN PNGDASVEED KQCKPNKKVI KTVPOLTTQD LKPETPENKV
151 DSTHOKHTHK PQPGVDHQS EKANEGREET DLEDEELMQ AYQCHVTEEM
201 AKEIKRKIRK KLKEQLTYFP SDTLFHDHKL SSEKRAKKKE VPVFSKAETS
251 TLTISGDTVE GEQKKESSVR SVSSDSHQDD EISSMEQSTE DSMQDDTKPK
301 PKTKKKKKA VADNNEDVDG DGVHEITSRD SPVYPKCLLD DDLVLGVYIH
351 RTDRLSKDFM ISHPMKIHV VDEHTGQYVK KDSGRPVSS YYEKENVDYI
401 LPIMTQPYDF KQLKSRLPEW EEQIVFNENF PYLLRGSDES PKVILFFEIL
451 DFLSVDEIKN NSEVQNEQCG FRKIWAFLK LLGANGNANI NSKLRLQLYY
501 PPTKPRSPLS VVEAFVWWSK CPRNHYPSTL YVTVRGLKVP DCIKPSYRSM
551 MALQEEKGKP VHCERHHESS SVDEPGLEE SKEVIKWRKL PGQACRIPNK
601 HFLFSNAGER GCFCLDFSHN GRILAAACAS RDGYPIILYE IPSGRFMREL
651 CGHLNIIYDL SWSKDDHYIL TSSSDGTARI WKNEINNTNT FRVLPHPSFV
701 YTAKFHFAVR ELVVTGCYDS MIRIWKVEMR EDSAILVRQF DVHKSFINSL
751 CFDTGEGHMY SGDCGTGVIV WNTYVKINDL EHSVHWTIN KEIKETEFKG
801 IPISYLEIHP NGKRLLIHTK DSTLRKINDL ILVARKFVGA ANVREKIHST
851 LTPCGTFLFA GSEDEIVYVW NPETGEQVAM YSDFPKSPI RDISYHPFEN
901 MVAFCARQGN EPILLYIYDF HVAQQAEMF KRYNGTFPLP GIHQSQDALC
951 TCPKLPHQGS FOIDEFVHTE SSSTKMQLVK QLETVTEVI RSCAAKVNKN
1001 LSFTSPPAVS SQSKLQKSN MLTAQEILHQ FGFTQTGIIS IERKPCNHQV
1051 DTAPTVALYV DYTANRSDEL TIHRGDIIRV FGDKNEDWYV GSIGKGQEGY
1101 PFANHVASET LYQELPPEIK ERSPLSPEE KTKIEKSPAP QKQSKNNKS
1151 QDFRLGSESM THSEMRKEQS HEDQGHIMDT RMRKNQAGR KVTLIE

```

## BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphtes3\_ln3, frame 1

TREMBL:U92792\_1 gene: "tup1"; product: "Tup1"; Schizosaccharomyces pombe general transcriptional repressor Tup1 (tup1) mRNA, complete cds., N = 1, Score = 186, P = 1e-10

TREMBL:AF104258\_1 gene: "Pmc733"; product: "putative copper-inducible 35.6 kDa protein"; Festuca rubra putative copper-inducible 35.6 kDa protein (Pmc733) mRNA, complete cds., N = 1, Score = 235, P = 4.6e-18

TREMBL:SPAC3H5\_8 gene: "SPAC3H5.08c"; product: "beta-transducin"; S.pombe chromosome I cosmid c3H5., N = 2, Score = 231, P = 2e-14

PIR:T02533 hypothetical protein F13M22.17 - Arabidopsis thaliana, N = 2, Score = 228, P = 1e-13

TREMBL:AF104258\_1 gene: "Pmc733"; product: "putative copper-inducible 35.6 kDa protein"; Festuca rubra putative copper-inducible 35.6 kDa protein (Pmc733) mRNA, complete cds., N = 1, Score = 235, P = 4.6e-18

TREMBL:SPAC3H5\_8 gene: "SPAC3H5.08c"; product: "beta-transducin"; S.pombe chromosome I cosmid c3H5., N = 2, Score = 231, P = 2e-14

TREMBL:CERO3E1\_1 gene: "R03E1.1"; Caenorhabditis elegans cosmid R03E1, N = 1, Score = 215, P = 2.3e-13

SWISSPROT:Y2LL\_CAEEL HYPOTHETICAL 43.1 KD TRP-ASP REPEATS CONTAINING PROTEIN K04G11.4 IN CHROMOSOME X., N = 1, Score = 203, P = 7.1e-13

>TREMBL:AF104258\_1 gene: "Pmc733"; product: "putative copper-inducible 35.6 kDa protein"; Festuca rubra putative copper-inducible 35.6 kDa protein (Pmc733) mRNA, complete cds.  
Length = 321

## HSPs:

Score = 235 (35.3 bits), Expect = 4.6e-18, P = 4.6e-18  
Identities = 59/225 (26%), Positives = 111/225 (49%)

```

Query: 647 MRELCGHLNIIYDLSWSKDDHYILTSSSDGTARIWKNEINNTNTFRVLPHPSFVYTAKFH 706
      + E GH + I DLSWSK+ +L++S D T R+W ++ + +V H ++V +P+
Sbjct: 63 VHEFYGHGDAILDLSWSKNGD--LLSASMDKTVRLW--QVGRDSCLKVFSHTNYVTCVOFN 119

Query: 707 PAVRELVVVTGCDYSMIRIWKVEMREDSAILVRQFDVHKSFINSCLFDTEGHMYSGDCTG 766
      +TGC D ++RIW V LV + K + ++C+ +G +G TG
Sbjct: 120 PTNGNYFITGCDGLRVINDVRK-----CLVVDWANSKEIVTAVCYRPDGGKAVAGTITG 174

Query: 767 VIVVMNTYVVKINDLEHSVHWTINKEIKETEFKGIPIISYLEIHPNGKRLLIHTKDSLRI 826
      ++ +LE V ++N K ++ Y P K+L++ + D+ +RI

```

Subjct: 175 NCRYYDASENRLEESQV---SLNGRKKSLHKRAIVGFQYCPSPDP--KKLMVTSGDAQVRI 229

Query: 827 MDLRILVARKEVGAANYREKINSTLTPCGTFLFAGSEGDIVYVWN 871

Subjct: 230 LDGAHVSN-YKGLQS-SSQVARSFTPDGDHIVSASDDSRIMWN 272

Pedant information for DKFZphtes3\_ln3, frame 1

Report for DKFZphtes3\_ln3.1

[LENGTH] 1196  
[MW] 137114.70  
[pI] 6.79  
[HOMOL] SWISSPROT:YKY4\_CAEEL HYPOTHETICAL 40.4 KD TRP-ASP REPEATS CONTAINING PROTEIN C14B1.4 IN CHROMOSOME III. 8e-21  
[FUNCAT] 99 unclassified proteins [S. cerevisiae, YKL121w] 2e-11  
[FUNCAT] 04.05.01.01 general transcription activities [S. cerevisiae, YBR198c TAF90 - TFIID subunit] 4e-10  
[FUNCAT] 30.10 nuclear organization [S. cerevisiae, YBR198c TAF90 - TFIID subunit] 4e-10  
[FUNCAT] 06.10 assembly of protein complexes [S. cerevisiae, YPR178w] 1e-08  
[FUNCAT] 04.05.03 mrna processing (splicing) [S. cerevisiae, YPR178w] 1e-08  
[FUNCAT] 03.22 cell cycle control and mitosis [S. cerevisiae, YDR364c] 4e-08  
[FUNCAT] 03.16 dna synthesis and replication [S. cerevisiae, YDR364c] 4e-08  
[FUNCAT] 08.07 vesicular transport (golgi network, etc.) [S. cerevisiae, YDL145c] 9e-08  
[FUNCAT] 30.09 organization of intracellular transport vesicles [S. cerevisiae, YDL145c] 9e-08  
[FUNCAT] 04.05.01.04 transcriptional control [S. cerevisiae, YCR084c] 2e-07  
[FUNCAT] 10.99 other signal-transduction activities [S. cerevisiae, YHL002w] 7e-07  
[FUNCAT] 98 classification not yet clear-cut [S. cerevisiae, YFR024c-a] 2e-06  
[FUNCAT] 02.16 fermentation [S. cerevisiae, YMR116c] 4e-06  
[FUNCAT] 30.03 organization of cytoplasm [S. cerevisiae, YMR116c] 4e-06  
[FUNCAT] 05.04 translation (initiation, elongation and termination) [S. cerevisiae, YMR116c] 4e-06  
[FUNCAT] 03.10 sporulation and germination [S. cerevisiae, YFL009w] 4e-05  
[FUNCAT] 03.04 budding, cell polarity and filament formation [S. cerevisiae, YFL009w] 4e-05  
[FUNCAT] 30.04 organization of cytoskeleton [S. cerevisiae, YFL009w] 4e-05  
[FUNCAT] 03.01 cell growth [S. cerevisiae, YCR088w] 6e-05  
[FUNCAT] 03.25 cytokinesis [S. cerevisiae, YCR057c] 7e-05  
[BLOCKS] BL00024H  
[SCOP] dltbgs\_2.46.3.1.1 betal-subunit of the signal-transducing 3e-91  
[SCOP] dlqfc\_2.21.2.1.9 Growth factor receptor-bound protein 2 (GRB2), N 4e-14  
[SCOP] dlrmk\_2.21.2.1.8 (1-64) c-src tyrosine kinase [human (Hom 5e-15  
[SCOP] dladd5b1\_2.21.2.1.7 (1-63) Hemopoietic cell kinase Hck [human (Hom 3e-15  
[SCOP] dlckal\_2.21.2.1.16 (1-54) p56-lck tyrosine kinase, SH3 domain [huma 1e-13  
[SCOP] dlqwea\_2.21.2.1.15 Src kinase, SH3 domain [Avian sarcoma virus 2e-15  
[SCOP] dlshg\_2.21.2.1.6 alpha-Spectrin, SH3 domain [chicken (Gallu 2e-13  
[SCOP] dlprmc\_2.21.2.1.13 Src kinase, SH3 domain [chicken (Gallus gallus) 2e-15  
[SCOP] dlhsq\_2.21.2.1.12 Phospholipase C, SH3 domain [human (Hom 2e-13  
[SCOP] dlboa\_2.21.2.1.3 Abl tyrosine kinase, SH3 domain [Mouse (Mu 3e-13  
[SCOP] dlafna\_2.21.2.1.2 Fyn, SH3 domain [human (Homo sapiens) 2e-15  
[SCOP] dlsema\_2.21.2.1.11 Growth factor receptor-bound protein 2 (GRB2), N 1e-13  
[SCOP] dlgbqa\_2.21.2.1.10 Growth factor receptor-bound protein 2 (GRB2), N 3e-16  
[SCOP] dlckaa\_2.21.2.1.1 C-Crk, N-terminal SH3 domain [mouse (Mu 3e-15  
[EC] 3.1.4.3 Phospholipase C 2e-07  
[EC] 3.1.4.11 1-Phosphatidylinositol-4,5-bisphosphate phosphodiesterase 7e-07  
[EC] 3.6.1.32 Myosin ATPase 7e-07  
[EC] 2.7.1.112 Protein-tyrosine kinase 8e-06  
[PIRKW] nucleus 2e-08  
[PIRKW] phosphotransferase 8e-06  
[PIRKW] plasma 4e-07  
[PIRKW] duplication 4e-07  
[PIRKW] phosphoric diester hydrolase 2e-07  
[PIRKW] tandem repeat 7e-07  
[PIRKW] hormone 4e-07  
[PIRKW] transmembrane protein 2e-06  
[PIRKW] stomach 4e-07  
[PIRKW] actin binding 7e-07  
[PIRKW] ATP 7e-07  
[PIRKW] phosphoprotein 7e-07  
[PIRKW] signal transduction 7e-09  
[PIRKW] heterotrimer 7e-09  
[PIRKW] P-loop 7e-07  
[PIRKW] hydrolase 7e-07  
[PIRKW] transcription regulation 5e-06  
[PIRKW] GTP binding 7e-09

[SUPFAM] 1-phosphatidylinositol-4,5-bisphosphate phosphodiesterase II 2e-07  
 [SUPFAM] SH3 homology 2e-07  
 [SUPFAM] SH2 homology 2e-07  
 [SUPFAM] protozoan myosin heavy chain IB 7e-07  
 [SUPFAM] myosin motor domain homology 7e-07  
 [SUPFAM] pleckstrin repeat homology 2e-07  
 [SUPFAM] protein-tyrosine kinase src 8e-06  
 [SUPFAM] WD repeat homology 3e-12  
 [SUPFAM] 1-phosphatidylinositol-4,5-bisphosphate phosphodiesterase domain Y homology 2e-07  
 [SUPFAM] protein kinase homology 8e-06  
 [SUPFAM] 1-phosphatidylinositol-4,5-bisphosphate phosphodiesterase domain X homology 2e-07  
 [SUPFAM] GTP-binding regulatory protein beta chain 7e-09  
 [SUPFAM] yeast coatomer complex alpha chain 4e-07  
 [PROSITE] RGD 1  
 [PROSITE] MYRISTYL 6  
 [PROSITE] AMIDATION 2  
 [PROSITE] CAMP\_PHOSPHO\_SITE 4  
 [PROSITE] CK2\_PHOSPHO\_SITE 25  
 [PROSITE] TYR\_PHOSPHO\_SITE 4  
 [PROSITE] PKC\_PHOSPHO\_SITE 19  
 [PROSITE] ASN\_GLYCOSYLATION 6  
 [PFAM] Src homology domain 3  
 [PFAM] WD domain, G-beta repeats  
 [KW] Irregular  
 [KW] 3D  
 [KW] LOW\_COMPLEXITY 5.77 %  
 [KW] COILED\_COIL 2.42 %

SEQ MPTAESEAKVTKVRFEKLLKTHSDLMREKKKKKKLVRSEENISPTIRSNLHYMKETT  
 SEG .....XXXXXXXXX.....  
 COILS .....CCCCCCCCCCCCCCCCCCCCCCCC.....  
 IgotB .....  
 SEQ SDDPDTIRSNLPHIKETTSDDVSAANTNLLKKSTRVTKNKLRLNTQLATENPNGDASVEED  
 SEG .....  
 COILS .....  
 IgotB .....  
 SEQ KQGKPNKKVIKTVPLTTQDLKPETPENKVDSTHQKTHTKPQGVHDQKSEKANEGREET  
 SEG .....XXXX.....  
 COILS .....  
 IgotB .....  
 SEQ DLEDEELMQAYQCHVTEEMAKEIKRKIRKKLKEQLTYFPSDTLPHDDKLSSEKRRKKKE  
 SEG .....XXXXXXXXXXXXXXXXXXXX.....  
 COILS .....  
 IgotB .....  
 SEQ VPFVSKAETSLTISGDTVEGEQKKESSVRVSSSDSHQDDEISSMEQSTEDSMQDDTKPK  
 SEG .....XXXXXXXXXXXX.....  
 COILS .....  
 IgotB .....  
 SEQ PKTKKKTKAVADNNEVDGVDGVEITSRDSVPYPKCLDDDLVLGVYIHRTRLKSDFM  
 SEG .....XXXXXXXX.....  
 COILS .....  
 IgotB .....  
 SEQ ISHPNVKIHVVDEHTGQYVKKDDSGRPVSSYYEKENVDYILPIMTPQYDFKQLKSRLPEW  
 SEG .....  
 COILS .....  
 IgotB .....  
 SEQ EEIVFNENFPYLLRGSDSPKVLFFIELDFLSVDEIKNNSEVQNECGGFKIAWAFK  
 SEG .....  
 COILS .....  
 IgotB .....  
 SEQ LLGANGNANINSKRLRLQLYPPPTKPSPLSVVEAFEWWSKCPRNHYPSTLYVTVRGLKVP  
 SEG .....  
 COILS .....  
 IgotB .....  
 SEQ DCIKPSYRSMALQEEKGKPVHCERHHSSSDTEPGLEESKEVIKWRLPQACRIPNK  
 SEG .....  
 COILS .....  
 IgotB .....

```

SEQ      HLFSLNAGERGCFCLDFSHNGRILAAACASRDGYPILLYEIPSGRFMRELCGHLNIIYDL
SEG      .....
COILS    .....
lgotB    .....CEEEEECCCCCEEE

SEQ      SWSKDDHYILTSSSDGTARIWKNEINNTNTFRVLPHPSFVYTAKFHPAVRELVTGCVDS
SEG      .....
COILS    .....
lgotB    EETTTTTEEEETTTEEEETT--TTCEEEETTTCCEEEETT--TCEEEETT

SEQ      MIRIWKVEMREDSAILVRQFDVHKSFINSLCFDTEGHHMYSGDCTGVIVVWNTYVKINDL
SEG      .....
COILS    .....
lgotB    EEEEEETTTTBTTEEEEEECCCCCE-EEEEETTEEEETTTEEEEEE.....

SEQ      EHSVHHWTINKEIKETEFKGIPIISYLEIHPNGKRLLIHTKDSTLRIMDLRILVARKFVGA
SEG      .....
COILS    .....
lgotB    .....

SEQ      ANYREKIHSTLTPCGTFLFAGSEDIYVYVNPETGEQVAMYSDFPKSPIRDISYHPFEN
SEG      .....
COILS    .....
lgotB    .....

SEQ      MVAFCAGQNEPILLYIDFHVAAQEAEMFKRYNGTFPLPGIHQSQDALCTCPKLPHQGS
SEG      .....
COILS    .....
lgotB    .....

SEQ      FOIDEFVHTESSTKMQLVKORLETVTEVIRSCAAKVNKNLSFTSPPAVSSQSKLKQSN
SEG      .....
COILS    .....
lgotB    .....

SEQ      MLTAQEIHQFGFTQTGTIISIERKPCNHQVDATPTVVALYDYTANRDELTHRGDIIRV
SEG      .....
COILS    .....
lgotB    .....

SEQ      FFKDNEDWYGSIGKGQEGYFPANHVASETLYQELPPEIKERSPLSPEETKIEKSPAP
SEG      .....
COILS    .....
lgotB    .....

SEQ      OKQSINKNKSQDFRLGSESMTHSEMRKEQSHEDQGHIMDTMRKNKQAGRKVTLIE
SEG      .....
COILS    .....
lgotB    .....

```

## Prosites for DKFZphtes3\_in3.1

PS00001	460->464	ASN_GLYCOSYLATION	PDOC00001
PS00001	686->690	ASN_GLYCOSYLATION	PDOC00001
PS00001	934->938	ASN_GLYCOSYLATION	PDOC00001
PS00001	1000->1004	ASN_GLYCOSYLATION	PDOC00001
PS00001	1065->1069	ASN_GLYCOSYLATION	PDOC00001
PS00001	1148->1152	ASN_GLYCOSYLATION	PDOC00001
PS00004	91->95	CAMP_PHOSPHO_SITE	PDOC00004
PS00004	264->268	CAMP_PHOSPHO_SITE	PDOC00004
PS00004	305->309	CAMP_PHOSPHO_SITE	PDOC00004
PS00004	1190->1194	CAMP_PHOSPHO_SITE	PDOC00004
PS00005	48->51	PKC_PHOSPHO_SITE	PDOC00005
PS00005	66->69	PKC_PHOSPHO_SITE	PDOC00005
PS00005	93->96	PKC_PHOSPHO_SITE	PDOC00005
PS00005	170->173	PKC_PHOSPHO_SITE	PDOC00005
PS00005	232->235	PKC_PHOSPHO_SITE	PDOC00005
PS00005	268->271	PKC_PHOSPHO_SITE	PDOC00005
PS00005	304->307	PKC_PHOSPHO_SITE	PDOC00005
PS00005	327->330	PKC_PHOSPHO_SITE	PDOC00005
PS00005	352->355	PKC_PHOSPHO_SITE	PDOC00005
PS00005	384->387	PKC_PHOSPHO_SITE	PDOC00005
PS00005	440->443	PKC_PHOSPHO_SITE	PDOC00005
PS00005	533->536	PKC_PHOSPHO_SITE	PDOC00005
PS00005	546->549	PKC_PHOSPHO_SITE	PDOC00005
PS00005	643->646	PKC_PHOSPHO_SITE	PDOC00005
PS00005	677->680	PKC_PHOSPHO_SITE	PDOC00005
PS00005	690->693	PKC_PHOSPHO_SITE	PDOC00005
PS00005	702->705	PKC_PHOSPHO_SITE	PDOC00005



PS00005	823->826	PKC_PHOSPHO_SITE	PDOC00005
PS00005	973->976	PKC_PHOSPHO_SITE	PDOC00005
PS00006	22->26	CK2_PHOSPHO_SITE	PDOC00006
PS00006	59->63	CK2_PHOSPHO_SITE	PDOC00006
PS00006	77->81	CK2_PHOSPHO_SITE	PDOC00006
PS00006	116->120	CK2_PHOSPHO_SITE	PDOC00006
PS00006	137->141	CK2_PHOSPHO_SITE	PDOC00006
PS00006	180->184	CK2_PHOSPHO_SITE	PDOC00006
PS00006	245->249	CK2_PHOSPHO_SITE	PDOC00006
PS00006	276->280	CK2_PHOSPHO_SITE	PDOC00006
PS00006	283->287	CK2_PHOSPHO_SITE	PDOC00006
PS00006	288->292	CK2_PHOSPHO_SITE	PDOC00006
PS00006	292->296	CK2_PHOSPHO_SITE	PDOC00006
PS00006	327->331	CK2_PHOSPHO_SITE	PDOC00006
PS00006	390->394	CK2_PHOSPHO_SITE	PDOC00006
PS00006	454->458	CK2_PHOSPHO_SITE	PDOC00006
PS00006	510->514	CK2_PHOSPHO_SITE	PDOC00006
PS00006	570->574	CK2_PHOSPHO_SITE	PDOC00006
PS00006	663->667	CK2_PHOSPHO_SITE	PDOC00006
PS00006	672->676	CK2_PHOSPHO_SITE	PDOC00006
PS00006	804->808	CK2_PHOSPHO_SITE	PDOC00006
PS00006	985->989	CK2_PHOSPHO_SITE	PDOC00006
PS00006	1023->1027	CK2_PHOSPHO_SITE	PDOC00006
PS00006	1127->1131	CK2_PHOSPHO_SITE	PDOC00006
PS00006	1132->1136	CK2_PHOSPHO_SITE	PDOC00006
PS00006	1161->1165	CK2_PHOSPHO_SITE	PDOC00006
PS00006	1170->1174	CK2_PHOSPHO_SITE	PDOC00006
PS00007	1083->1091	TYR_PHOSPHO_SITE	PDOC00007
PS00007	211->219	TYR_PHOSPHO_SITE	PDOC00007
PS00007	1083->1091	TYR_PHOSPHO_SITE	PDOC00007
PS00007	210->219	TYR_PHOSPHO_SITE	PDOC00007
PS00008	483->489	MYRISTYL	PDOC00008
PS00008	577->583	MYRISTYL	PDOC00008
PS00008	716->722	MYRISTYL	PDOC00008
PS00008	800->806	MYRISTYL	PDOC00008
PS00008	861->867	MYRISTYL	PDOC00008
PS00008	941->947	MYRISTYL	PDOC00008
PS00009	811->815	AMIDATION	PDOC00009
PS00009	1188->1192	AMIDATION	PDOC00009
PS00016	1074->1077	RGD	PDOC00016

## Pfam for DKFZphtes3\_ln3.1

HMM_NAME	WD domain, G-beta repeats		
HMM	*MrCHnnWVMCVaFSPDGrWFIvSGSWDgTCRLWD*		
Query	650	LCGHLNIIYDLWSKDDHY-ILTSSSDGTARIWK	682
HMM_NAME	Src homology domain 3		
HMM	*pyVIALYDYqAqdpDELSFkEGDIIiIEsDD.WWrgRnnnTNGQEGW		
Query	1054	PTVVALYDYTANRSDELTIHRGDIIRVFFKDNEDWWYGSIGK--GQEGY	1100
HMM	IPSNYVEPi*		
Query	1101	FPANHVASE	1109

DKFZphtes3\_20c21

-----

group: testes derived

DKFZphtes3\_20c21 encodes a novel 708 amino acid protein without similarity to known proteins.

No informative BLAST results; No predictive prosite, pfam or SCOP motif.

The new protein can find application in studying the expression profile of testis-specific genes.

unknown

Sequenced by MediGenomix

Locus: /map="22q11.2-12.2"

Insert length: 3997 bp

Poly A stretch at pos. 3877, polyadenylation signal at pos. 3853

```
1 GGTAGGCGGG GCGGCGCGTG ACCTAAGGCC TCTCTGCCGC GCGCGCAGGT
51 ACGGGGCAGA AGTCGCAAGT ACCCAGCTGC TGCCACGTT TCTGGTCCAG
101 AGTCCCGAAC CCGAGCACT GGGATGCTG GCTACTCGA GCCAAGGCAC
151 TGATGTTTGA ACTGGAACCT TCAAAACGTT TAATAAGAGT CTTCAGGATG
201 GGTTTGAAC TACAAAGCTA GAAATTTCTT TAGAACACCA GCTCTAGCAT
251 GCATCTCCCA CTTTTGCTT TCCTGGAGAG GAGCTTGAAG AGGTGGTCTT
301 GCAGACAGCC ACAGTGATAC TCAGGAAACC AGAGGAATGG ATTTGACTTT
351 TCTGCTAGGA TTCTTTGTTA TAGTTTCTCC CTGAGTTGTA AGAGGCATGG
401 AAATATACAT GAAACTGAAG AACCTGCAAG GAAGGGAAGT GGAACCTTCC
451 ATGCTGAGTG AAAACTAACC AAGTGGCAGT TGTGACTGAA AACACTGAAA
501 CCTACCAAGT CCAGATTAC TGGATTGGGG GATAGAGGAA CGGTACACAG
551 TAGGAGAGAA GAAGTGATAC CGGAAAAGAA AACCTAAATG AAGAGAATGA
601 GGATGACTGC ACAGTAGATG GCCACCTCTA CCTCCACAGA GGCAAGTCA
651 GCCTCGTGGT GGAATTAATT TTTCTTTAT GATGGTTCGA AGSTAAAGGA
701 AGAAGGCGAT CCAACAAGAG CTGGCATTTC TTACTTTTAT CCTTCCAGA
751 CCCTGCTAGA CCAACAGGAG TTGCTTTGTC GACAGATTGC TGGAGTTGTC
801 CGCTGTCTTT CTGACATTTT TGACTCTGCT CACTACTCTG TTGCTCTGAG
851 AAAACGAGAG TTTGCCATAA AAGTTGATGG AGATTACCTT TGGGTCTGGG
901 GCTGTGCTGT GGAGCTCCCT GATGTCACT GCAAGCGGTT TCTGGATCAG
951 CTAGTTGGAT TCTTTAATT TTACAATGGA CCTGTTTCCC TAGCTTATGA
1001 GAATGTTTCT CAGGAAGAAG TGAGCAGCGA GTGGGACACC TTCATCGAGC
1051 AAATCTGTA AAACACCACT GATCTGCATA AGATTTTCAA TTCCCTCTGG
1101 AACTTGGACC AAATAAAGT GGAGCCCTCT TTGTTGCTGA AGGCAGCCCG
1151 CATTCTGAG ACCTGCCAGC GCTGCCCTCA CATTCTGCTT GGCTGCATCC
1201 TCTATAAAGC ACTGATTGTC AGCACCCAAC TCCCGCCCTC CCTCACCCGG
1251 AAGGTCCTGC TTCACCGAAC AGCACCTCAG GAGCAGAGAC TCCCTACGGG
1301 AGGGGATGCC CCGCAGGAAC ATGGAGCGGC ATTGCCCCCG AATGTCCAGA
1351 TTATCCCTGT TTTTGTGACC AAAGAGGAAG CCATTAGTCT CCACGAGTTC
1401 CCGGTGGAAC AGATGACAAG GTCTCTAGCA TCTCCAGCAG GACTCCAGGA
1451 TGGTTACGCC CAGCACCATC CAAAGGGTGG GAGCACATCT GCCCTGAAG
1501 AAAACGCCAC TGCCCATGTG GAATCCATGG CTTGGACAC CCAGATCCC
1551 ACATCCCTCT ACGAAGCTTG TCCAGATGGC AGGAAGGAGA ACGGATGCTT
1601 GTCTGGCCAT GATCTGGAGA GCATCAGGCC CGCAGGACTG CACAACCTG
1651 CAGGGGTGA GGTCTTGGC CTACGCTCTT CCCTGGGAA GGAACCTAGT
1701 TTTCTCAAG AAGAACTCGA CTGTCTGAA ATCCACATTC CAGAGGCTCA
1751 GGAAGTGAAG ATGGCTCAG GTCAATTTGC CTTCCTACAT GTGCCCTGTC
1801 CAGATGGCAG GGCTCCTTAC TGCAAGGCAT CTCTCAGCGC TCCAGCAGC
1851 CTGGAACCCA CGCTCTCTGA GGACACAGCC ATCAGCAGCT TGCGCCCTCC
1901 CTCTGCTCTT GAGATGCTGA CCGCATGAG AGCCCAAGAG CAGGTCGAAG
1951 ACCATCTCTG CCATAGCAGC CAAGCCCTCA TTCCAGAGC AGACCTCTCT
2001 CCCAGAAGGA CCGCAGGCC CTGTTATTG CTTGCTTAG ATCCAGGACA
2051 GAGAGAAAC AAGCTTCCCA CGGGGAAAC AGGCCTGGAT GAGGATGTTG
2101 ATGGGGTCTG TGAAGCCAC GCAGCCCTCT GTCTGGAATG CAGTTCAAGC
2151 TCAGCAAACT GTCAGGCTGC TGCCCTCTCT GCAGATGGAA TCAGCTCCAG
2201 GCTGACACCA GCAGAGTCTT GCATGGGGCT CGTGAGGATG AATCTCTACA
2251 CTCACCTGCT CAAAGGGCTG ATGCTGTCCC TGCTGCTGTA GAGGCCCTG
2301 CTGGAGACA CCGCAGCCTT AGAGGAAGT TACCACACCA GCCTGGCTTC
2351 ACTGAATGGG CTGGAAGTCC ACCTGAAGA GACGCTGCCC AGGGATGAGG
2401 CAGCTCCAC GAGCAGCACC TACAACCTCA CATATTACGA CCGCATTCAG
2451 AGCTTGTGA TGGCAACCT GCCGAGGTG GCCACCCGCG ATGATCGCCG
2501 CTCTCTCAG GCGTCAGCC TGATGCATAG CGAATTTGCC CAGCTGCCCG
2551 CGCTTTATGA AATGACTGTC AGAATGCTT CCACGGCTGT GTACGCCTGT
2601 TGCAACCCCA TCCAGGAGC ATATTTCCAG CAGCTGGCAC CTGCAGCAGC
2651 GAGCTCCGCG TTCCCAAAAC CTCAGGATGG CGCCTTCAGC CTCTCCGGCA
2701 AAGCAAGACA GAAGTCTGCT AAGCAGCGGG TGAATTTGCT CTGAACCTGA
2751 CCCAGGAGGT GACTGGGAAG GAGAAAACCA GCAAGGAAG CTCTGCCTTT
2801 TATAATTGAA AAGGCCCTCT TATTTTATTT TTCTGAAA CATTCCCTTT
```

```
2851 TTTAGGAACC AAATGATATT TGAGTTTTTG TTATTCCTTT TGCAGATTGG
2901 GATGTGTTTT GGGGGCAGGG GTTAGTCTTT CAGGTCGGCA GACCCAGAGC
2951 ACTTGATAAA GAACTGTATT TAATCGGTAG TGTGGGGGCC GGGACGGGCT
3001 TGGCTCCCTC TCGCCATAC TGAGCCTGAG GTATTTCATA TCTCCTGCTG
3051 TTCCATCCCA GCTTGAATTG GTGCCACAAG CTTCGAAGTT GGCATTTTTT
3101 CTAGAACCCTG ATCGTCCACT AGCCCAGAGT GTGTGTGTTT AACCCCCACA
3151 CCAGGTGGTG GTAGGCGGTG TGACTGCACA GCGAGGTGCC GGATCTGTGA
3201 GCAGGCCGAC TCCACTCCCA CGCCGAGGT AGGTTTCTCC AGTGGGCTCT
3251 TGCTGGGAGG TCCGGATCGT TCCTGCAGGG AAGCGGCAGC ACACGGAGAC
3301 CACTGGTTG AATTCTGTG GAACTCTACT CAAATCTAGG GGCCTCTTCT
3351 TTGGACCCAC AATGGGGGCA AGCCTTAATA ATATGGAAGG GAGTTTGGGC
3401 TTTAGAGATC CCTTTATAAA AGCTCTGGGG GCTGAGCCCT GAGAATTCAG
3451 TGACAACAGG ACCAACCCTG GCTGCTTTTG ACTCAAGTG GGCCTGCAG
3501 CTGGTTCTCT TCGAGCGAGT GTCCCTAAAT AGGAGTTTAC AAGATGCTTG
3551 GGGTAAAG CACTGTGCTT TTCAGTGGTG GCTGCGTAA AGGAGCGGAC
3601 ACTCAGCTGT GTGTTCTGG GCTTGTGTG TACTTAGAAC CTCAGTTCTA
3651 TTACGTTATA CTCAGACATT TTTTTCAGC TATGAGACAG ACTGCAGGAT
3701 GAAATATTT GTCAAAATCT TAACGAATG TTTACTGGAA GTACTTGAGA
3751 TTCCATTGGA GAGTTGTATT GTTAATAAT TCATGTCAGT GAACTGATAT
3801 CTGATGTTTA TGATATGGTG TCTTTTCTT GAAACAAGCT TCCAAGGGCT
3851 AGAAATAAAA TAGCCAAAAA ATGCTGAAA AAAAAAAAAA AAAAAAAAAA
3901 AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA
3951 AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAA
```

## BLAST Results

Entry HS1048E9 from database EMBLNEW:  
Human DNA sequence from clone 1048E9 on chromosome 22q11.2-12.2  
Contains pseudogene similar to ribosomal protein S3A and part of a gene  
similar to C.elegans protein CE02118, ESTs, STS, GSS.  
Score = 6540, P = 0.0e+00, identities = 1308/1308  
-14 exons

## Medline entries

No Medline entry

## Peptide information for frame 3

ORF from 618 bp to 2741 bp; peptide length: 708  
Category: putative protein  
Classification: no clue

```
1 MATSTSTEAK SASWNYEFL YDGSKVKEEG DPTRAGICYF YPSOTLLDQO
51 ELLCCQIACV VRCVSDISDS PPTLVRLRLK KFAIKVDKDY LNWLGCAVEL
101 PDVSCRFELD QLVGFNFYN GPVSLAYENC SQEELSTEMD TFIEQLKNT
151 SDLHKIFNSL WNLQTKVEP LLLLKAARIL QTCORSFIL AGCILYGLI
201 VSTQLPPLSL AKVLLHRTAP QEORLPTGGD APOEHGAALP PNVQIIPFV
251 TKEEAILHE FPVEQMTRSL ASPAGLQDGS AQHHPKGGST SALKENATGH
301 VESMAWTTPO PTPDEACPD GRKENGCLSG HDLESIRPAG LHNSARGEVL
351 GLSSSLGKEL VFLQEELDLS EIHIPAEQEV EMASGHFAFL HVPVPDGRAP
401 YCKASLSASS SLEPTPPEDT AISSLRPPSA PEMLTQMGAG EQVEDHPGHS
451 SQAPIPRADP LPRRTRRPLL LPRLDPGQRG NKLPTGEQGL DEDVDGVCE
501 HAAPGLECSS GSANCQAGP SADGISSRLT PAESCHMLVR MNLYTHCVKG
551 LMLSLLAEEP LLGDSAAIEE VYHSSLASLN GLEVHLKETL PRDEAASTSS
601 TYNFTYYDRI QSLMANLPQ VATPHDRRL QAVSLMHSEF AOLPALYEMT
651 VRNASTAVYA CCNPIQETYE QQLAPARRSS GFPNPDGAF SLSGKAKQKL
701 LKHGVNLL
```

## BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphtes3\_20c21, frame 3

No Alert BLASTP hits found

Pedant information for DKFZphtes3\_20c21, frame 3

Report for DKFZphtes3\_20c21.3

[LENGTH] 708  
[MW] 76900.23  
[pI] 5.30  
[KW] Alpha\_Beta  
[KW] LOW\_COMPLEXITY 6.36 %

```
SEQ MATSTSTEAKSASWNNYFFLYDGSKVKEEGDPTRAGICYFYPSQTLDDQQLCGQIAGV
SEG .XXXXXXXXXXXXX.....
PRD ccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc
```

```
SEQ VRCVSDISDSPPTLVRLRLKFAIKVDGYLWVLGCAVELPDVSKRFLDQLVGFFNFYN
SEG .....
PRD eeeeecccccccccccccccccccccccccccccccccccccccccccccccccccccc
```

```
SEQ GPVSLAYENCQEELSTEWDTFIEQILKNTSDLHKIFNSLWNLDTKVEPLLLKKAARIL
SEG .....
PRD ccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc
```

```
SEQ QTCQRSFILAGCILYKGLIVSTQLPPSLTAKVLLHRTAPQEQRLPTGGDAPQEHGAALP
SEG .....
PRD hhhcccccccccccccccccccccccccccccccccccccccccccccccccccccccccc
```

```
SEQ PNQIIPVFTKEEATSLHEFPVEQMTSLASPAQLDQSAQHHPKGGSTSALENATGH
SEG .....
PRD ccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc
```

```
SEQ VESMAWTTDPDTPSPDEACPDGRKENGCLSGHDLESIRPAGLHNSARGEVLGLSSSLGKEL
SEG .....
PRD ccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc
```

```
SEQ VFLQEELDLSEIHIEAQEVEMASGHFAFLHVPVPGRAPYCKASLSASSSLEPTPPEDT
SEG .....
PRD hhhhhhcccccccccccccccccccccccccccccccccccccccccccccccccccccccccc
```

```
SEQ AISSLRPPSAPEMLTQGAQEVEDHPGHSSQAPIPRADPLPRRTRPLLLPRLOPGRG
SEG .....
PRD ccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc
```

```
SEQ NKLPTGEQGLDEDVDGVCESHAAPGLECSSGSANCQAGPSADGISSRLTPAESCMGLVR
SEG .....
PRD ccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc
```

```
SEQ MNLYTHCVKGLMLSLLAEPLLGDSAITEEVYHSSLASLNGLEVHLKETLPRDEAASTSS
SEG .....
PRD ceeeeehhhhhhhhhcccccccccccccccccccccccccccccccccccccccccc
```

```
SEQ TYNFTYYDRIQSLMANLPQVATPHDRRFLQAVSLMHSEFAQLPALYEMTVRNASTAVYA
SEG .....
PRD ccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc
```

```
SEQ CCNPIQETFFQQLAPARSSGFNPQDGAFLSGKAKQKLLKHGVNLL
SEG .....
PRD ecccccccccccccccccccccccccccccccccccccccccccccccccccccccccc
```

(No Prosite data available for DKF2phtes3\_20c21.3)

(No Pfam data available for DKF2phtes3\_20c21.3)

DKFZphtes3\_20k2

group: signal transduction

DKFZphtes3\_20k2 encodes a novel 839 amino acid protein with strong similarity to rat vanilloid receptor subtype 1.

VR1 seems to play an important role in the activation and sensitization of nociceptors. It is the receptor for e.g. capsaicin, a selective activator of nociceptors, a natural product of capsicum peppers. The novel protein is the human orthologue of rat VR1.

The new protein can find application as a target for the development of new nociception-modulating drugs.

strong similarity to rat vanilloid receptor subtype 1

Sequenced by MediGenomix

Locus: unknown

Insert length: 4187 bp

Poly A stretch at pos. 4154, polyadenylation signal at pos. 4135

```
1 GGCTCAGGCA GGCCTGGCCC AGAGTCACGC TGGCAACCAC GAGTTTGGGA
51 AGCAGTCGTA TTCTCTCTCT CTCTCTCTCT CTCTCAGTAT CCATGACAGT
101 GTGATGGAGA GTCTCTGCGG TGCCATCTGG GATCCAAACC GTCCCTGTGT
151 CCCCACGTC CAGGCCCTAG ATGCTCCCGC CCGGTCACTC ACTTACTCGT
201 CAGATCGGCC GTCTCTGTAT CACAGTCTCT CTGTTCAAGT TGCACACTGG
251 GCCACAGAGG ATCCAGCAAG GATGAAGAAA TGGAGCAGCA CAGACTTGGG
301 GGCAGCTGGC GACCCATCCG AAAAGGACAC CTGCCAGAC CCCTTGGATG
351 GAGACCCCTAA CTCAGGCCCA CTCAGGCCA AGCCCCAGCT CTCACGGGCC
401 AAGAGCCGCA CCGGCTCTTT TGGGAAGGCT GACTCGGAGT AGGCTTTCCT
451 GGTGGATTGC CCTCAGGAGG AAGGTGAGCT GGAATCTGTC CCGACCATCA
501 CAGTCAGCCC TGTATCACC ATCCAGAGGC CAGGAGACGG CCCACCCGGT
551 GCCAGGCTGC TGTCCTCAGG CTCTGTCTGC GCCAGCACCG AGAAGACCTT
601 CAGGCTCTAT GATCGCAGGA GTATCTTTGA AGCCCTTCTT CAGAATAACT
651 GCCAGGATCT CGAGAGCCCTG CTGCTCTTCC TGCAGAGAGC CAAGAAGCAC
701 CTCACAGACA AGCAGTTCAA AGACCTCTAG ACAGGGAAGA CTTCTCTGCT
751 GAAAGCCATG CTCACCTGTC ATGACGGACA GAACACCACC ATCCCTCTGC
801 TCCTGGAGAT CGCGCGGCAA ACGGACAGCC TGAAGGAGCT TGTCAACGCC
851 AGCTACAGGC ACAGTACTTA CAAGGGCCAG ACAGCAGTGC ACATCGCCAT
901 CGAGAGAGGC AACATGGCCC TGGTGACCTT CCTGGTGGAG AACGGAGCAG
951 ACGTCCAGGC TGGCGCCCAT GGGGACTTCT TTAAGAAAAC CAAAGGGCGG
1001 CCTGGATTCT ACTTCGGTGA ACTGCCCTTG TCCTGGCCGG CGTGACCAA
1051 CCAGCTGGGC ATCTGGAAGT TCCTGCTGCA GAACTCTCTG CAGACGGCCG
1101 ACATCAGCGC CAGGACTCTG GTGGGCAACA CGGTGCTGCA CGCCCTGGTG
1151 GAGGTGGCCG ACAACACGGC CGACAACACG AAGTTTGTGA CGAGCATGTA
1201 CAATGAGATT CTGATCTCTG GGGCCAAACT GCACCCGAGC CTGAAGCTGG
1251 AGAGCTCTAC CAACAAGAAG GGAATGACGC CGCTGGCTCT GGCAGCTGGG
1301 ACCGGGAAGA TCGGGGTCTT GGCCTATATT CTCACGGGGG AGATCCAGGA
1351 GCCCGAGTGC AGGCACCTGT CCAGGAAGTT CACCGAGTGG GCCTACGGGC
1401 CCGTGCACTC CTGCTGTGAC GACCTGTCTT GCATCGACAC CTGCGAGAAG
1451 AACTCGGTGC TGGAGGTGAT CGCCTACAGC AGCAGCGAGA CCCCTAATCG
1501 CCACGACATG CTCTTGTGGG AGCCGCTGAA CCGACTCTCT CAGGACAAGT
1551 GGGACAGATT CGTCAAGCGC ATCTTCTACT TCAACTTCTT GGTCTACTGC
1601 CTGTACATGA TCATCTTCAC CATGGCTGCC TACTACAGGC CCGTGGATGG
1651 CTTGCTCTCC TTTAAGATGG AAAAAATTGG AGACTATTTT CGAGTTACTG
1701 GAGAGATCCT GTCTGTGTTA GGAGGAGTCT ACTTCTTTTT CGAGGGGATT
1751 CAGTATTTCG TGCAGAGCGG GCGCTGATG AGACCCCTGT TTGTGGACAG
1801 CTACAGTGAG ATGCTTTTCT TTCTGCAGTC ACTGTTCTATG CTGGCCACCG
1851 TGGTGTCTGA CTTCAGCCAC CTCAGGAGAT ATGTGGCTCT CATGTTATTC
1901 TCCTTGGCCT TGGGCTGGAC CAACATGCTC TACTACACCC GCGGTTTCCA
1951 GCAGATGGGC ATCTATGCCG TCATGATAGA GAAGATGATC CTGAGAGACC
2001 TGTGCCGTTT CATGTTGTG TACATCTCTT TCTGTTTCGG GTTTTCCACA
2051 GCGGTGGTGA CGCTGATTGA AGACGGGAAG AATGACTCCC TGCCGTCTGA
2101 GTCCACGTGC CACAGGTGGC GGGGGCTGCT CTGAGGCCCC CCCGATAGCT
2151 CCTACAACAG CCTGTACTCC ACCTGCTTGG AGCTGTTCAA GTTCACCATC
2201 GGCATGGGCG ACCTGGAGTT CACTGAGAAC TATGACTTCA AGGCTGTCTT
2251 CATCATCCCTG CTGCTGGCCT ATGTAATTCT CACCTACATC CTCCTGCTCA
2301 ACATGCTCAT CGCCCTCATG GGTGAGACTG TCAACAAGAT CGCAGAGGAG
2351 ACCAGAACCA TCTGGAGCTC GCAGAGAGCT ATCACCATCC TGGACACGGA
2401 GAAGAGCTTC CTTAAGTGCA TGAGGAAGGC CTTCCGCTCA GGCAGCTGTC
2451 TGCAGGTGGG GTACACACCT GATGGCAAGG ACGACTACCG GTGGTGCTTC
2501 AGGGTGGACG AGGTGAACCT GACCACCTGG AACACCACCG TGGGATCAT
2551 CAACGAAGAC CCGGGCAACT GTGAGGGCGT CAAGCGCACC CTGAGCTTCT
2601 CCCTGCGGTC AAGCAGAGTT TCAGGCAGAC ACTGGAAGAA CTTTGCCCTG
2651 GTCCCCCTTT TAAGAGAGGC AAGTGCTCGA GATAGGCAGT CTGCTCAGCC
```

```
2701 CGAGGAAGTT TATCTGCGAC AGTTTTTCAGG GTCTCTGAAG CCAGAGGACG
2751 CTGAGGTCTT CAAGAGTCCT GCCGCTTCCG GGGAGAAGTG AGGACGTCAC
2801 GCAGACAGCA CTGTCAACAC TGGGCCTTAG GAGACCCCGT TGCCACGGGG
2851 GGCTGCTGAG GGAACACCAG TGCTCTGTCA GCAGCCTGGC CTGGTCTGTG
2901 CCTGCCCAGC ATGTCTCCAA ATCTGTGCTG GACAAGCTGT GGGAAAGCGT
2951 CTGGGAAGCA TGGGAAGTGA TGTACATCA ACCGTCACTG TCCCCAAGTG
3001 AATCTCCTAA CAGACTTTCA GCTTTTACT CACTTTACTA AACAGTTTGG
3051 ATGTGCAGTC TCTACTGGGA CATGTTAGGC CCTTGTTC TTTGATTTTA
3101 TTCTTTTTTT TGAGACAGAA TTTCACCTTT CTCACCCAGG CTGGAATGCA
3151 GTGGCACAAT TTGGCTCCC TGCAACCTCC GCCTCTGGA TTCCAGCAAT
3201 TCTCTGCGCT CGGCTTCCCA AGTAGCTGGG ATTACAGGCA CGTGCCACCA
3251 TGTCTGGCTA ATTTTTGTA TTTTTTAAT AGATATGGGG TTTCGCCATG
3301 TTGGCCAGGC TGGTCTCGAA CTCCTGACCT CAGGTGATCC GCCCACCTCG
3351 GCCTCCCAAA GTGCTGGGAT TACAGGTGTG AGCCTCCACA CCTGGCTGTT
3401 TTCTTTGATT TTATCTTTT TTTTTTTCT GTGAGACAGA GTTCACTCT
3451 TGTTCGCCAG GCTGGAGTGC AGTGGTGTGA TCTTGGCTCA CTGCAACCTC
3501 TGCTTCCCGG GTTCAAGCGA TTCTTCTGCT TCAGTCTCCC AAGTAGCTTG
3551 GATTACAGCT GAGCACTACC ACCGCCGCT AATTTTGTG TTTTAAAG
3601 AGACGGGGTT TCACCATGTT GGCCAGGCTG GTCTCGAACT CTGACCTCA
3651 GGTGATCTGC CCGCCTTGGC CTCCCAAGT GCTGGGATTA CAGGTGTGAG
3701 CCGCTGCGCT CGGCCTTCTT TGATTTTATA TTATTAGGAG CAAAAGTAAA
3751 TGAAGCCAG GAAACACCTT TTGGGAACAA ACTCTCTCTT TGATGGAAAA
3801 TGCAGAGGCC CTCTCTCTCT GTGCGTGTCT TGCTCTCTTT ACCTGCCCGG
3851 GTGGTTTGGG GGTGTTGGTG TTCTCTCCTT GGAGAAGATG GGGGAGGCTG
3901 TCCCACTCCC AGCTTGGCA GAATCAAGCT GTTGCGAGCA TGCTTCTTTC
3951 ATCTCTCTT ACGATCAATC ACAGTCTCCA GAAGATCAGC TCAATTGCTG
4001 TGCAGGTAA AACTACAGAA CCACATCCCA AAGGTACCTG GTAAGAATGT
4051 TTGAAGATC TTCCATTTCT AGGAACCCCA GTCTGTGCTT TCCGCAATGG
4101 CACATGCTTC CACTCGATC ATACTGGCAT CCTCAAAATA ACAGATATGT
4151 ATACATATAA AAAAAAAAAA AAAAAAAAAA AAAAAA
```

## BLAST Results

No BLAST result

## Medline entries

99288727:  
Recent advances in neuropharmacology of cutaneous nociceptors.

99231880:  
A non-pungent triphenyl phenol of fungal origin, scutigerol, stimulates rat dorsal root ganglion neurons via interaction at vanilloid receptors.

## Peptide information for frame 2

ORF from 272 bp to 2788 bp; peptide length: 839  
Category: strong similarity to known protein  
Classification: Cell signaling/communication

```
1 MKKWSSTDLG AAADPLQKDT CPDPLQGDPN SRPPFAKPOL STAKSRTRLF
51 KGDSEEAFF VDCPHEGEL DSCPTITVSP VITIQRFGDG PTGARLLSQD
101 SVAATEKTL RLYDRRSIFE AVAQWQDOL ESLLLFLQKS KKHLDNEFK
151 DPETGKCLL KAMLNLDHGO NTIPLLEI ARQDLSLKEC VNASYDSYY
201 KGQTAHIAI EERNMALVTL LVENGADVOA AAHGDFFKKT KGRPGFYFGE
251 LPLSLAECTN QLGIKVLFLQ NSWQADISA RDSVGNVTLH ALVEVADNTA
301 DNTKFVTSY NEILILGAKL HPTLKLELT NKKGMTPLAL AAGTGKIGVL
351 AYILQREIQE PECHLSRKF TEWAYGVPVS SLYDLSCIDT CEKNSVLEVI
401 AYSSSETPNR HDMLLVEPLN RLLQDKWDRF VKRIFYFNFL VYCLYMIIFT
451 NAAYYRFPVG LPPFKMEKIG DYFRVTGEIL SVLGGVYFFF RGIQYFLQRR
501 PSMKTLFVDS YSEMLFLOS LEMLATVVLY FSHLKEYVAS MVFSALQWT
551 NMLYYTRGFQ QMGIYAVMIE KMILRDLCRF MFVYIVFLFG FSTAVVTLIE
601 DGNDSLPSSE STSHRWGPA CRPPSSSYNS LYSTCLELFF FTIGMDLEF
651 TENYDFKAVF IILLAYVIL TYILLNMLI ALMGETVNNI AQESKNIWKL
701 QRAITLQTE KSLKCHMKA FRSGKLQVG YTPDCKDDYR WCFRVDENVW
751 TWWTWVIGII NEDPGCEGV KRTLSFLSRS SRVSCRHWKN FALVPLLREA
801 SARDRQSAQP EEVYLRQFSG SLKPDAEVF KSPAASCEK
```

## BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphtes3\_20k2, frame 2

TREMBL:AF029310\_1 product: "vanilloid receptor subtype 1"; Rattus norvegicus vanilloid receptor subtype 1 mRNA, complete cds., N = 1, Score = 3760, P = 0

TREMBLNEW:AB015231\_1 product: "stretch-inhitable nonselective channel (SIC)"; Rattus norvegicus mRNA for stretch-inhitable nonselective channel (SIC), complete cds., N = 2, Score = 2090, P = 2e-219

>TREMBL:AF029310\_1 product: "vanilloid receptor subtype 1"; Rattus norvegicus vanilloid receptor subtype 1 mRNA, complete cds.  
Length = 838

#### HSPs:

Score = 3760 (564.1 bits), Expect = 0.0e+00, P = 0.0e+00  
Identities = 721/839 (85%), Positives = 773/839 (92%)

```
Query: 1 MKKWSSTDGAAADPLQKDTCPDPLDGPNSRPPPAKPQLSTAKSRTLFGKGDSEEAFF 60
M++ +S D + P Q+++C DP D DPN +PPP KP + T +SRTLFGKGDSEEA P
Sbjct: 1 MEQRASLSESESESPQENSCLOPPORDPNCKPPVPKHIFITRSRTLFGKGDSEEAFF 60

Query: 61 VDCPHEEGELDSCTITVSPVITIQRPDGPFGTARLLSQDSVAASSTKTLRLYDRRSIFE 120
+DCP+EEG L SCP ITVS V+TIQRPDGP R SQDSVA EK RLYDRRSIFE
Sbjct: 61 LDCPYEEGGLASCPITVSSVLTIQRPDGPASVRPSSQDSVSAG-EKPPRLYDRRSIFE 119

Query: 121 AVAQNNQCQLESLLFLQKSKKHLTDNEFKDPETGKTKLLKAMLNLDGQNTTIPLLLEI 180
AVAQ+NCQ+LESLL FLQ+SKK LTD+EFKDPETGKTKLLKAMLNLDH+GQN TI LLL++
Sbjct: 120 AVAQSNQCQLESLLFLQKSKKRLTDEFEKDPETGKTKLLKAMLNLDHNGQNTIALLLDV 179

Query: 181 ARQTDLSKELVNASYTDSSYKGTALHIAIERRNMLVTLVENGADVQAAAHGDFFKKT 240
AR+TDSLK+ VNASYTDSSYKGTALHIAIERRNM LVTLVENGADVQAAA+GDFFKKT
Sbjct: 180 ARKTDLSKQFVNASYTDSSYKGTALHIAIERRNMLVTLVENGADVQAAAHGDFFKKT 239

Query: 241 KGRPGFYFGELPLSLAACTNQLGIVKFLQNSWOTADISARDSVGNTVLHALVEADNTA 300
KGRPGFYFGELPLSLAACTNQL IVKFLQNSWQ ADISARDSVGNTVLHALVEADNT
Sbjct: 240 KGRPGFYFGELPLSLAACTNQLAIVKFLQNSWQPADISARDSVGNTVLHALVEADNTV 299

Query: 301 DNTKFVTSMYNEILILGAKLHPTLKLEELTNKKGMTPLAALAGTGKIGVLAYILQREIQE 360
DNTKFVTSMYNEILILGAKLHPTLKLEE+TN+KG+TPLALAA +GKIGVLAYILQREI E
Sbjct: 300 DNTKFVTSMYNEILILGAKLHPTLKLEELTNKRGKTPPLAALASSGKIGVLAYILQREIHE 359

Query: 361 PECRHLSRKFTWAYGPHVSSLYDLSCIDTCEKNSVLEVIAYSSSETPNRHMDLVEPLN 420
PECRHLSRKFTWAYGPHVSSLYDLSCIDTCEKNSVLEVIAYSSSETPNRHMDLVEPLN
Sbjct: 360 PECRHLSRKFTWAYGPHVSSLYDLSCIDTCEKNSVLEVIAYSSSETPNRHMDLVEPLN 419

Query: 421 RLLQOKWDRFVKRIFYFNFLVYCLYMIIFTMAAYRPVQGLPPFKMEK-IGDYFRVTGEI 479
RLLQOKWDRFVKRIFYFNFLVYCLYMIIFTMAAYRPVQGLPPFKMEK-IGDYFRVTGEI
Sbjct: 420 RLLQOKWDRFVKRIFYFNFLVYCLYMIIFTMAAYRPVQGLPPFKMEK-IGDYFRVTGEI 479

Query: 480 LSVLGGVYFFFRGIQYFLQRRPSMKTLFVDSYSEMLFFLQSLFMLATVVLYFSHLKEYVA 539
LSV GGVYFFFRGIQYFLQRRPS+K+LFVDSYSE+LFF+QSLFML +VVLYFS KEYVA
Sbjct: 480 LSVSGGVYFFFRGIQYFLQRRPSLKSLEFVDSYSEILFFVQSLFMLVSVVLYFSQRKEYVA 539

Query: 540 SMVFSLALGWTNMLYYTRGFQMGIIYAVMIEKMLRDLCRFMFVYVFLFGFSTAVVTLI 599
SMVFSLA+GWTNMLYYTRGFQMGIIYAVMIEKMLRDLCRFMFVY+VFLFGFSTAVVTLI
Sbjct: 540 SMVFSLAMGWTNMLYYTRGFQMGIIYAVMIEKMLRDLCRFMFVYVFLFGFSTAVVTLI 599

Query: 600 EDGKNDSLPESTSHRWGPACRPDSSYNSLYSTCLELFKFTIGMGDLFTENYDFKAV 659
EDGKN+SLP EST H+ RG AC+P +SYNSLYSTCLELFKFTIGMGDLFTENYDFKAV
Sbjct: 600 EDGKNSLPESTPHKRGSAKCP-GNSYNSLYSTCLELFKFTIGMGDLFTENYDFKAV 658

Query: 660 FIILLAYVILTYILLNMLIALMGETVNKIAQESKNWKLQRAITILOTEKSFCLKMRK 719
FIILLAYVILTYILLNMLIALMGETVNKIAQESKNWKLQRAITILOTEKSFCLKMRK
Sbjct: 659 FIILLAYVILTYILLNMLIALMGETVNKIAQESKNWKLQRAITILOTEKSFCLKMRK 718

Query: 720 AFRSGKLLQVGYTPDGKDDYRMCFRVDEVNWTWNTNVGIINEDPGNCEGVKRTLSFSLR 779
AFRSGKLLQVGYTPDGKDDYRMCFRVDEVNWTWNTNVGIINEDPGNCEGVKRTLSFSLR
Sbjct: 719 AFRSGKLLQVGYTPDGKDDYRMCFRVDEVNWTWNTNVGIINEDPGNCEGVKRTLSFSLR 778

Query: 780 SSRVSGRHWKFNALVPLLRASARDROSAQPEEVYLRQPSGSLKPEDAEEVFKSPAASGEK 839
S RVSGR+WKFNALVPLLR+AS RDR + Q EEV L+ ++GSLKPEDAEEVFK GEK
Sbjct: 779 SGRVSGRHWKFNALVPLLRASARDRQAEVQVQKHYTGSLKPEDAEEVFKSDSMVPEK 838
```

Pedant information for DKFZphtes3\_20k2, frame 2

## Report for DKF2phtes3\_20k2.2

[LENGTH] 839  
[MW] 94950.75  
[PI] 6.90  
[HOMOL] TREMBL:AF029310.1 product: "vanilloid receptor subtype 1"; Rattus norvegicus  
vanilloid receptor subtype 1 mRNA, complete cds. 0.0  
[FUNCAT] 99 unclassified proteins [S. cerevisiae, YIL112w] 4e-05  
[PIRKW] alternative splicing 3e-06  
[PIRKW] peripheral membrane protein 3e-06  
[SUPFAM] ankyrin repeat homology 3e-06  
[SUPFAM] unassigned ankyrin repeat proteins 3e-06  
[PFAM] Ank repeat  
[KW] TRANSMEMBRANE 4

SEQ MKKWSSTDGAAADPLQKDTCPDPLDGDPSRPPPAKPOLSTAKSRTLFGKGDSEEAFF  
PRD ccc  
MEM .....  
SEQ VDCPHEEGELDSCPTITVSPVITIQPGDGTGARLLSQDSVAATEKTLRLYDRRSIFE  
PRD ccc  
MEM .....  
SEQ AVAQNQCQDLESLLFLQSKKHLTDNEFKDPETGKTCLLKAMLNLHDGQNTTIPLLLEI  
PRD hhhhechhhhhhhhhhhhhhhcccccccccccccccccccccccccccccccccccc  
MEM .....  
SEQ ARQTDSELKELVNASYTDYKGTALHIAIERRNMALVTLLVENGADVQAAAHGDFKKT  
PRD hhhcc  
MEM .....  
SEQ KGRPGFYFGEPLSLAACTNQLGIVKFLQNSWQTADISARSDVGNVTVLHALVEADNTA  
PRD ccc  
MEM .....  
SEQ DNTKFVTSMYNEILILGARLHPTLKEELTNKKGMTPLAAGTGKIGVLAVIILQREIQE  
PRD chhhhhhhhhhhhhhhcc  
MEM .....  
SEQ PECRHLRSKFTWAYGPHSSLYDLSDICTEKNVLEVIAYSSSETFNRHDMLLVEPLN  
PRD ccc  
MEM .....  
SEQ RLLQDKWDRFVKRIFYFNFVLYCLYMIIFTMAAYRVDGLPPFKMEKIGDYFRVTGEIL  
PRD hhh  
MEM .....  
SEQ SVLGGVYFFFRGIQYFLQRRPSMKTLEVDYSEMLFFLOSLFMLATVVLYFSLKEYVAS  
PRD ccc  
MEM .....  
SEQ MVFSLAGWTNMLYYTRGFQMGYAVMIEKMLRLDLCRFMFVYIVFLGFSTAVVTLE  
PRD hhh  
MEM .....  
SEQ DGKNDLPSESTSHRWGPFACRPDSSYNSLYSTCLELFTKFTIGMGDLEFFENYDFKAVF  
PRD ccc  
MEM .....  
SEQ IILLAYVILTYILLNMLIALMGETVKNIAQESKNIWKLQRAITILDTEKSLFKCMRKA  
PRD hhh  
MEM .....  
SEQ FRSGKLLQVGYTPDGKDDYRWCFRVDEVMTTNTNTNVIINEDPGNCEGVKRTLFSLSRS  
PRD hhccc  
MEM .....  
SEQ SRVSGRHWKNFALVPLLREASARDRQSAQPEEVYLRQFSGSLKPEDAIEVFKSPAASGEK  
PRD ccc  
MEM .....

(No Prosite data available for DKF2phtes3\_20k2.2)

Pfam for DKF2phtes3\_20k2.2



WO 01/12659

PCT/IB00/01496

HMM_NAME	Ank repeat
HMM	*GyTPLHIAARYNNVEMVrLLQHGADIN*
	G+T+LHIA +++N+ +V LL+++GAD+
Query	202 GQTALHIAIERNNMALVTLLVENGADVQ 229

DKF2phtes3\_2013

group: transmembrane protein

DKF2phtes3\_2013 encodes a novel 595 amino acid protein with partial similarity to the IL-17 receptor.

The novel protein contains one transmembrane region.  
No informative BLAST results; No predictive prosite, pfam or SCOP motif.

The new protein can find application in studying the expression profile of testis-specific genes and as a new marker for testicular cells.

similarity to IL-17 receptor

Sequenced by MediGenomix

Locus: unknown

Insert length: 2406 bp

Poly A stretch at pos. 2345, no polyadenylation signal found

```
1 GCCTCAGGTG TTCTGCGTT GTTGTGAGT GGAGAGCAGG GAGTGGGGCC
51 AGCCAGCAGA AACAGTGGGC TGTACACAT CACCTTCAA TATGACAATT
101 GTACACACCA CTGGAATCCA GTGGGGAAGC ATGTGATTGC TGACGCCACG
151 AATATCACCA TCAGCCAGTA TGCTTGCCAT GACCAAGTGG CAGTCACCAT
201 TCTTTGGTCC CCAGGGGCCC TCGGCATCGA ATTCTTGAAA GGAATTCGGG
251 TAATACGGGA GGAGCTGAAG TCGGAGGGA GACAGTCCA ACAACTGATT
301 CTAAGGATC CGAAGCAGCT CAACAGTAGC TTCAAAAGAA CTGGAATGGA
351 ATCTCAACCT TTCTGAATA TGAATTTGA AACGGATTAT TTCTGAAGG
401 TTGTCCCTTT TCCTTCCATT AAAACGAAA GCAATTACCA CCCTTCTTC
451 TTTAGAACCC GAGCCTGTGA CCTGTTGTTA CAGCCGACCA ATCTAGCTTG
501 TAAACCCCTC TGGAAGCCTC GGAACCTGAA CATCAGCCAG CATGGCTCGG
551 ACATGCCAGT GTCCCTCGAC CAGCACCAGC ACAACTCTCG CTTCCTTTTC
601 TTCTATCTTC ACTACAAGCT CAAGCACGAA GGACCTTTCA AGCGAAAGAC
651 CTGTAAAGCAG GAGCAAACTA CAGAGATGAC CAGCTGCCTC CTTCAAAATG
701 TTTCTCCAGG GGAATTATATA ATTGAGCTGG TCGATGACAC TAACACAACA
751 AGAAAGATGA TCCATTATGC CTTAAAGCCA GTGCACCTCC CGTGGCCGGG
801 GCCCATCAGA GCCGTGGCCA TCACAGTGCC ACTGGTAGTC ATATCGGCAT
851 TCGCGACGCT CTTCACTGTG ATGTGCCGCA AGAAGCAACA AGAAATATA
901 TATTCACATT TAGATGAAGA GAGCTCTGAG TCTTCCACAT ACACCTGCAGC
951 ACTCCCAAGA GAGAGGCTCC GGCCGCGGCC GAAGGTCTTT CTCTGCTATT
1001 CCAGTAAAGA TGGCCAGAAT CACATGAATG TCGTCCAGTG TTTGCGCTAC
1051 TTCTCCAGG ACTTCTGTGG CTGTGAGGTG GCTCTGGACC TGTGGGAAGA
1101 CTTGAGCCTC TGTAGAGAAG GGCAGAGAGA ATGGGTCTATC CAGAAGATCC
1151 ACGAGTCCCA GTTCATCATT GTGGTTTGTT CCAAGGTTAT GAAGTACTTT
1201 GTGGACAAGA AGAATACAA ACACAAAGGA GGTGGCCAGC GCTCGGGGAA
1251 AGGAGAGCTC TTCTTGTGGG CGGTGTGAGC CATTCGCCAA AACCTCCGCC
1301 AGGCCAAGCA GAGTTCTCTC GCGGCGCTCA GCAAGTTTAT CGCCGCTCTAC
1351 TTTGATTATT CTTGCGAGGG AGACGTCCCC GGTATCTTAG ACCTGAGTAC
1401 CAAGTACAGA CTCATGAGCA ATCTTCTCTA GCTCTGTTC CACTGCACT
1451 CCCGAGACCA CGGCTCTCAG GAGCCGGGCC AGCACACGGC ACAGGGCAGC
1501 AGAAGGAACCT ACTTCCGGAG CAAGTCAGGC CGGTCCCTAT ACCTGCCCAT
1551 TTGCAACATG CACCAAGTTA TTGACGAGGA GCCCGACTGG TTGAAAGAGC
1601 AGTTCTGTTCC CTTCCATCCT CCTCCACTGC GCTACCGGGA GCCAGTCTTG
1651 GAGAAATTGG ATTCCGGCTT GGTTTTAAAT GATGTCNTGT GCAAAACAGG
1701 CCTGAGAGGT GACTTCTGCC TAAAGTAGA GCGGCTCTT CTTGGGGCAA
1751 CCGGACACAG CGACTCCAG CACGAGAGTC AGCATGGGGG CTTGGACCAA
1801 GACGGGAGG CCGGCGCTGC CTTGACGGT AGCGCCGCCG TGCAACCCCT
1851 GCTGCACACG GTGAAAGCCG GCAGCCCTC GGACATGCGG CGGAGCTCAG
1901 GCATCTATGA CTCGTCTGTG CCCTCATCCG AGCTGTCTCT GCCACTGATG
1951 GAAGGACTCT CGACGGACCA GACAGAAACG TCTTCCCTGA CGGAGAGCGT
2001 GTCTCTCTCT TCAGGCTCGG GTGAGGAGGA ACCTCTCTCC CTTCCTTCCA
2051 AGCTCTCTCT TTCTGGGTCA TGCAAGCAG ATCTTGGTGG CCGCAGCTAC
2101 ACTGATGAAC TCCACGGCGT CGCCCTTTTG TAACAAAACG AAAGAGTCTA
2151 AGCATTGCCA CTTTAGCTGC TGCCCTCTC TGAATGCCCA GCTCATCTCC
2201 CTGGTTGCCAT GGGCCAGCTG GAGCTGAGGT CTATACAGG GATATTTCGA
2251 GTGAATGCTT GGGCAGTACT TGTCTCCCT TGCCCAACG CTTTACCGGA
2301 TATCTTGACA AACTCTCCAA TTTTCTAAA TGATATGGAG CTCTGAAAAA
2351 AAAAAA AAAAAA AAAAAA AAAAAA AAAAAA
2401 AAAAAA
```

## BLAST Results

No BLAST result

## Medline entries

No Medline entry

## Peptide information for frame 1

ORF from 346 bp to 2130 bp; peptide length: 595  
Category: similarity to known protein  
Classification: unclassified

```
1 MESQPLNMK FETDYFKVV PFPSIKNES YHPFFFRTRA CDLLQPDNL
51 ACKPFWKPRN LNISQHGSDM QVSFDHAPHN FGFRFFYLHY KKHGEPFKR
101 KTKQEQTTE MTSCLLNQVS PGDYIILVD DTNTRKVMH YALKPVHSPW
151 AGPIRAVAIT VPLVVISAFSA TLFTVMCRKK QQENIYSHLD EESSESSTYT
201 AALPRERLRP RPKVFLCYSS KDGQNMHNVV QCFAYFLQDF CGCEVALDLW
251 EDFSLCREGO REWVIQKINE SQFIIVVCSK GMRIFVDKRN YKHKGGGRGS
301 GKGEFLFVAV SAIAEKLRQA RQSSSAAISR FIAYVFDYSC EGDVPGILD
351 STKYRLMDNL POLCSHLHRS DHGLQEPQCH TROGSRNRYF RSKSGRSLYV
401 AICNMHQFID EEPDFEKFQF VPFHPPPLRY REPVLEKFDG GLVLNDVMCK
451 PGPESDFLCK VEAALVGATG PADSQHESQH GGLDQDGEAR PALDGSAAIQ
501 PLLHTVKAGS PSDMPRDSGI YDSSVPSEL SLPLMEGLST DQTETSSLTE
551 SVSSSSGLGE EEPALPSKL LSSGSKADL GCRSYTDELH AVAPL
```

## BLASTP hits

No BLASTP hits available

## Alert BLASTP hits for DKFZphtes3\_2013, frame 1

TREMBL:U58917.1 product: "IL-17 receptor"; Homo sapiens IL-17 receptor  
mRNA, complete cds., N = 1, Score = 215, P = 4.7e-14

TREMBL:NM31993.1 product: "interleukin 17 receptor"; Mus musculus  
interleukin 17 receptor mRNA, complete cds., N = 2, Score = 152, P =  
1.1e-13

>TREMBL:U58917.1 product: "IL-17 receptor"; Homo sapiens IL-17 receptor  
mRNA, complete cds.  
Length = 866

## HSPs:

Score = 215 (32.3 bits), Expect = 4.7e-14, P = 4.7e-14  
Identities = 85/284 (29%), Positives = 131/284 (46%)

```
Query: 213 KVFLCYSSKDGQNMHNVVQCFAYFLQDFCGCEVALDLWEDFSLCREGQREWV-IQK---I 268
KV++ YS+ D +++VV FA FL CG EVALDL E++ G WV QK +
Sbjct: 379 KVMIIYSA-DHPLYVDVVLKFAQFLLTACGTEVALDLLEEQAISEAGVMTWVGRQKQEMV 437

Query: 269 HESQFIIVVCSKGMKY---FVDKKNYXXXXXXXXXXELFLVAVSAIAEXXXXXXXXX 324
+ IIV+CS+G + + + +LF A++ I
Sbjct: 438 ESNKIIIVLCRGRTRAKWQALLGRGAPVRLRCDHGKFPVGLFTAAHMHILPDFKRPACFG 497

Query: 325 XXXXXFIAYVF-DYSCGDPVPGILDSTKYRLMDNLPLCSHLHSDHGLQEPQHTRQ 383
++ YF + SC+GDPV + + +Y LMD ++ + +D + +PG+ R
Sbjct: 498 T-----YVVCYFSEVSCDGTVPDLFGAAPRYPLMDRFEEV--YFRIQDLEMFQGRMHRV 550

Query: 384 G--SRNRYFRSKGRSLYVAICNMHQFIDEEPWFEEKQFV----PFHPPPLR---YREP 434
G S NY RS GR L A+ + PDWFE + + P L + EP+
Sbjct: 551 GELSGDNYLRSPGGRQLRAALDRFRDWQVRCPDWFECENLYSADDDQDAPSLDEEVFEEL 610

Query: 435 LEKFDGLVLNDVMCKPGPESDFLCKVEAALVGATGPADSQHESQHGLDQDGEAR 491
L +G+V + + P S CL++ V G A H L G+ P
Sbjct: 611 LPP-GTIVKRAPLVRE-PGSQACLAIDPLV-GEEGAIVAKLEPH--LQPRGQPAP 662
```

## Pedant information for DKFZphtes3\_2013, frame 1

Report for DKFZphtes3\_2013.1

[LENGTH] 595  
[MW] 66847.05  
[pI] 6.27  
[HOMOL] TREMBL:MM31993\_1 product: "interleukin 17 receptor"; Mus musculus interleukin  
17 receptor mRNA, complete cds. 2e-14  
[BLOCKS] BL00740A MAM domain proteins  
[BLOCKS] BL01224B N-acetyl-gamma-glutamyl-phosphate reductase proteins  
[KW] TRANSMEMBRANE 1  
[KW] LOW\_COMPLEXITY 13.61 %

SEQ MESQPFNLNMFETDYFVKVVPFSPKSNESNYHPFFRTRACDLLLPDNLACKPFWKPRN  
SEG .....  
PRD ccc  
MEM .....

SEQ LNISQHGSDMQVSFDHAPHNFGFRFFYLHYKLKHEGPFKRKTKQEQTTEMTSCLLQNV  
SEG .....  
PRD eeeeecc  
MEM .....

SEQ PGDYIIELVDDTNTTRKVMHYALKPVHSPWAGPIRAVAITVPLVVISAFATLFTVMCRK  
SEG .....  
PRD ccc  
MEM .....MMMMMMMMMMMMMM.....

SEQ QQENIYSHLDEESSESSTYTAALPRERLRPRPKVFLCYSSKDGQNHMNVVQCFAFLQDF  
SEG .....xxxxxxx.....xxxxxxx.....  
PRD hhhhhhhhhcc  
MEM .....

SEQ CGCEVALDLWEDFSLCREGQREWVIQKIHESQFIIIVCSKGMKYFVDKKNYKHKGGGRGS  
SEG .....xxxxxxx.....  
PRD ccchhhhhhhcc  
MEM .....

SEQ GKGELFLVAVSAIAEKLRAKQSSSAALSKEFIIVFDYSCGDPGILDSTKYRLMDNL  
SEG .....xxxxxxxxxxxxxxxx.....  
PRD cccccceehhh  
MEM .....

SEQ PQLCSHLHSRDHGLQEPGQHTRQGSRRNYFRSKSGRSLYVAICNMHQFIDEEDWFEKQF  
SEG .....  
PRD cchhhhhhhcc  
MEM .....

SEQ VPFHPPPLRYREPVLKFDGLVLDVMCKPGPESDFCLKVEAAVLGATGPADSQHESQH  
SEG .....  
PRD ecc  
MEM .....

SEQ GGLDQDGEARPALDGSAAQLPLHTVKAGSPSDMPRDSGIYDSSVPSSLSLPLMEGLST  
SEG .....xxxxxxxxxxxxxxxx.....  
PRD ccc  
MEM .....

SEQ DQTETSSLTESVSSSSGLGEEPPALPSKLLSSGSKADLGCRSYTDELHAVA  
SEG .....xxxxxxxxxxxxxxxx.....  
PRD hhhhhhhhhhecc  
MEM .....

(No Prosite data available for DKFZphtes3\_2013.1)

(No Pfam data available for DKFZphtes3\_2013.1)

DKFZphtes3\_20m18

group: nucleic acid management

DKFZphtes3\_20m18 encodes a novel 132 amino acid protein with similarity to the *S. cerevisiae* mitochondrial carrier protein RIM2.

The novel protein contains a leucine zipper and a Prosite mitochondrial energy transfer proteins signature. It is member of a family of substrate carrier proteins which are found in the inner mitochondrial membrane and are involved in energy transfer. The RIM2/MRS12 gene encodes a predicted protein of 377 amino acids that is essential for mitochondrial DNA metabolism and proper cell growth. Inactivation of this gene causes the total loss of mitochondrial DNA and, compared to wild-type rho<sup>0</sup> controls, a slow-growth phenotype on media containing glucose. The novel protein seems to be the human orthologue of this protein.

The new protein can find application in modulation of mitochondrial DNA replication and maintenance.

similarity to carrier protein RIM2

Sequenced by MediGenomix

Locus: unknown

Insert length: 3572 bp

Poly A stretch at pos. 3530, polyadenylation signal at pos. 3510

```
1 GCGCGGGGA GGGCTGTGCC GGTGCTTTC TGCAGCCGCA TCTCGGCCAG
51 CTCTCCTCGC CGTCCCCGGG GCGCTGTGCG TCTCCAGTCC GGGACCGAAG
101 CCGCCTGCCG TAGCGGGCGG CCAGATCCGC GTCCCGCCTC AGCGGCCGGA
151 GGACATGCCG GAGAGAGAAT GAGCCAGAGG GACACGCTGG TGCATCTGTT
201 TGCCGGAGGA TGTGGTGGTA CAGTGGGAGC TATTCTGACA TGCCACTGG
251 AAGTTGTAAA AACACGACTG CAGTCATCTT CTGTGACGCT TTATATTCT
301 GAAGTTCAGC TGAACACCAT GGCTGGAGCC AGTGTCAACC GAGTAGTGTC
351 TCCCGGACCT CTTCAATTGCC TAAAGTGAT CTGGAAAAA GAAGGGCCTC
401 GTTCCTTGTT TAGAGGACTA GGCCCCAATT TAGTGGGGGT AGCCCCTTCC
451 AGAGCAATAT ACTTTGCTGC TTATTCAAAC TGCAAGGAAA AGTTGAATGA
501 TGTATTGAT CCTGATTCTA CCCAAGTACA TATGATTTCA GCTGCAATGG
551 CAGGTATGAA TGTATAATAT TAAAAAATAA AAAAATTTC TGAACCTAG
601 AGGCTTAATA TTGAATTATA AGTTTGTAGT GAAAAGTTGA TGATTAATGT
651 GCTTTTTCATT GATTAGATGA TTTTACGTT TATCGATATA AACCAATTA
701 GGTATATGTA AAATCTGTCA TCAGTTGACA TTTTGTAGT CAGGAGTTA
751 CATGCTAGGG TACAAGTAAT ATATTATAT TGCCTTGTGT AGTCCACTGA
801 ATGTTTAGTG ATCATTGTGA ACAGTTTAA GAATCCAACC ATAATTACAC
851 TATAAATAAG TTATGGAGCT GTAATTTACT CTCTCTCCT CAATTTCTGT
901 TAGTGCCTTT TCCCTTTTGT CTGCATGTTT TGGCTTCTGT CTGAAATGTG
951 TCGGCAATTC TTGGTAAAGT ATTCATTTTG TCCTGTGCTC AAATGCTGAA
1001 ATTTTGTGTA GTGATGTATT ATTATTGACA ATTCAGTTAC TATGTGTATT
1051 TTTTAAATTT GTTTATTATT CTACATAATT CACACTAGAC AGCACCTGAA
1101 ATTTAGACAC TGGCTATGTG TACATGCTTA CTATAGAAAT GTTTCAGGA
1151 ACTCTCTGTT TCTGTCATCA CTGATAAGTA TATATGATTG TGAATTAATA
1201 TAACTAGTTT TAGGCTTTTA CCCTGCCATA AAGATAAACA GTTGGTTTGA
1251 CCAATCTGGT TCTGGAATCA TTTGCTGCTA TGCATGTTAG ACAAGCCAC
1301 GAACCTTGAT TTTCCATTGA AAATCTCCTC TAATATCTGA GATTTATTGT
1351 ATATTACTCT ATATCTCACA TTTTCAAAT ATGCTGTAAC TTTATAAACT
1401 GTAGCTGCTT TCATCAGCTA TTGATCAATA AATTGAATGT CAATTATGTG
1451 CTTAATAATG AGTGCCTTAA ACTGTTAAAC ACTTTTGGTT TAGAAATAAA
1501 GTGAATCAAT TTGACCTATA TACTTCATGA AGTAAGTAAG TTTGAAATAC
1551 AAATTTCTGA AAGTCAATA GCCCTTATCG TATTACAAAT TGTTTTAAAG
1601 GCTTTTGTGA TTTATTAATT GTCAGTTGAT TCACTGAAGC TTTAAACTG
1651 GAAGGGACAA TCCAAGGTC AAAAGAGTGA AATACAATCA TTTACCAATA
1701 AGGAAACCTT GGGCAAAATTA TGTAATTTAT GTGAACCTCT CTTAGCTTAC
1751 CCATGGAATG AGTCAAGTGG TCTACATAGA TTTGGATTTT GAGAATTAGT
1801 TCTTTTCATT AGTGTATATG AGATTATCTT GTTACAACCTA GAATTATTTT
1851 TAATGTAATT TTTACAGATG TTGAATATTA GTAGATAGGA TTTTCCCTC
1901 ACGAATTTGG ATGTAAGGTA AAGGTTGGTG GCCAGTGACA AACCTTATAA
1951 CCACCTTTATC AGGTTCTTTA AAAATATATT TGTGAATTAC CAGTGATTAT
2001 GTTTTGGGCT TATAACCTCA GATAATTATA AAGAAATGTT AATCTTATTT
2051 GAAAGAATTG GAATCTAGAA AGTTAGATGA GCAGTCATTT TATATTGATA
2101 TTTGTTATAT CAGTATAGCA AATGCAGAGG TTCAGAATAT CTTTATTTCC
2151 ACTGGAACAT CTTATTTCTT TAGAGTATCT CATCAGAATT TATTACTGTA
2201 TTTGTATCAC ATTGCAAAGA ATTTCAGTAG AATTGTCAGT TTGCACTTTT
2251 TTCTCAAATG TGTACAAATG TTAACATATA GTTCATTTTT ATCTGTACAT
2301 TGATGCCAAT TCCCAACTTG AATTCCTCAA GTTTTGGTAA ACTTACAATC
2351 TCATACTTGT TCAGAGGTTA TTGCACTGTA CACTTACTGT GTAGAAAAAT
2401 CTGTTTGAAT TTGTTGTCAG TTACATTGTT CTGAGAACTG TGCTCTCAGA
2451 GCTTCTGTGC ACTATTCATG AGCATTAACA CTTAGCCTTG CAGTTTTATA
```

```

2501 CATAACTATA TGGTTAGTAA AACTGAATGG TCCAATGCAG ACTCATTAAA
2551 GTAGGCTTTT GCCCCCTTTG TTCTTGAAAT AATCTAGACC AGATTACTCG
2601 GGGTTTTTTT TAGGATTATT TTTATAGGTC TAAATATGAA TGATTTGGGG
2651 GTATGAAGTA CTTAAAGATA GTTCTGTGAA AAATCATTTT CAGCTGTCTA
2701 TTCAAGGGAA AAAATGCTAA CCTTGTCAC TTTACTACACA AAACCACT
2751 AAAATAAACC ATTAATGATA CTGCCTGCAA GATTTTAAAC CACCAGATAG
2801 CACACACATT AAGGATTAT AAGGCACTGT ACGTAATTT TATCCAAGT
2851 GACCTCTCAA TTCATTTTCA TTTTGCATT TATCCATATG AACTCATGTT
2901 TAATTTAGAT AATAAAAATT TATTTTATTA AAAGGACAGT TTATTTAAAG
2951 TGGGTCCTTT TATTGTGTGT AGTGCATACT ATAAGAATTT GTAAGCCTCT
3001 AAAGTTGAGC TATAAATTTT CATGCATTAA AAATTTGTTT CAGTTGTGAG
3051 GATATTTAAT CAGATTAAAT AATGTTGACT CTTAATATTT TGCTTGCTT
3101 TTTTCTCTCC TACACATGAC CTTTGACAGA CTAAGTATAT CTCAGCTATT
3151 GAGGGTATCT GTTTTGTGTC CTGTATATTT TGTTTAAATT AACTTGTATA
3201 TTCTTTGTA TACACCTAGG CACAGATGTA TGCAAAAAAA ATTTGTTAAA
3251 TTACTTCTTT CTTTATACTA ATTCTCAATT TTTAAAAGAT TTTATCTGGC
3301 ATGTATATAC TTTTATATAG AACATTATAA ATGTAAAGGA AATGAATCT
3351 AATTTTAAAT GGATTATGTA TTCATACAGT TATCTCAAT TTTTAAATA
3401 CTAATAATGT AATCATTGAA TGTTTCCTAC ATACGTAGTG GGTTTTATTT
3451 GCTCACAGCA TACAGTTATT TTTCAATTTA TGTTTTCTA TTAGACTTAA
3501 ATTCATTAT AATAAAGGCT TTTACTCATT AAATACAAAA AAAAAA
3551 AAAAAA AA

```

## BLAST Results

No BLAST result

## Medline entries

No Medline entry

95198680:

Overexpression of a novel member of the mitochondrial carrier family rescues defects in both DNA and RNA metabolism in yeast mitochondria.

## Peptide information for frame 1

ORF from 169 bp to 564 bp; peptide length: 132  
 Category: similarity to known protein  
 Classification: Intracellular transport and traffic  
 Prosite motifs: LEUCINE\_ZIPPER (27-49)  
 MITOCH\_CARRIER (26-36)

```

1 MSQRDLVHL FAGGCGGTVG AILTCPLEVV KTRLQSSSVT LYISEVQLNT
51 MAGASVNRV SPGPLHCLKV ILEKEGPRSL FRGLGPNLVG VAPSRAYFA
101 AYSNCKEKLN DVFDPDSTQV HMISAAMAGM NV

```

## BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFzphes3\_20ml8, frame 1

PIR:S44092 probable carrier protein c2 - *Caenorhabditis elegans*, N = 2,  
 Score = 147, P = 1.5e-19

PIR:S36081 probable carrier protein RIM2, mitochondrial - yeast  
 (*Saccharomyces cerevisiae*), N = 1, Score = 230, P = 6.2e-19

>PIR:S36081 probable carrier protein RIM2, mitochondrial - yeast  
 (*Saccharomyces cerevisiae*)  
 Length = 377

## HSPs:

Score = 230 (34.5 bits), Expect = 6.2e-19, P = 6.2e-19  
 Identities = 55/133 (41%), Positives = 80/133 (60%)

```

Query:      8 VHLFAGGCGGTGAILTCPLEVVKTRLQSSS-VTLYISEVQLNTMAGA----SVNRVSP 62
              VH AGG GG GA++TCP ++VKTRLQS + Y S+ +N G+ S+N V+
Sbjct:     54 VHFVAGGIGGMAGAVVTCFDFLVKTRLQSDIFLKAYSQA-VNISKGSTRPKSINYVIQA 112

```

Query: 63 GP-----LHCLKVILEKEGPRSLFRGLGPNLVGVAPSRAIFYAAYSNCCKLNDVFD--P 115  
 G L + + ++EG RSLF+GLGPNLVGV P+R+I F Y K+ F+  
 Sbjct: 113 GTHFKETLGIIGNVYKQEGFRSLFKGLGPNLVGVIPARSINFFTYGTTKDMYAKAFNNGQ 172

Query: 116 DSTQVHMISAAMAG 129  
 ++ +H+++AA AG  
 Sbjct: 173 ETPMIHLMAAATAG 186

Score = 77 (11.6 bits), Expect = 1.1e+00, P = 6.8e-01  
 Identities = 25/88 (28%), Positives = 39/88 (44%)

Query: 3 QRDTLVHLFAGGCGGTVGAILTCPLEVVKTRLQSSSVTLYISEVQLNTMAGASVNRVVP 62  
 Q ++HL A G A T P+ ++KTR VQL+ SV + +  
 Sbjct: 172 QETPMIHLMAAATAGWATATATNPILIKTR-----VQLDKAGKTSVRQYKNS 219

Query: 63 GPLHCLKVILEKEGPRSLFRGLGPNLVG 90  
 CLK ++ EG L++GL + +G  
 Sbjct: 220 WD--CLKSVIRNEGFTGLYKLSASYLG 245

Score = 71 (10.7 bits), Expect = 6.6e+00, P = 1.0e+00  
 Identities = 28/91 (30%), Positives = 45/91 (49%)

Query: 12 AGGCGGTVGAILTCPLEVVKTRLQSSSVTLYISEVQLNTMAGASVNRVVPGLHCLKVI 71  
 + G V +I T P EVV+TRL+ + + N G R + G + KVI  
 Sbjct: 294 SAGLAKFVASIATYPHEVVRTRLRQTP-----KEN---G---KRKYT-GLVQSFKVI 338

Query: 72 LEKEGPRSLFRGLGPNLVGVAPSRAIFYAAY 102  
 +++EG S++ GL P+L+ P+ I F +  
 Sbjct: 339 IKEEGLFSMYSGLTPLHMRTPVNSIIMFGTW 369

Pedant information for DKFZphtes3\_20m18, frame 1

#### Report for DKFZphtes3\_20m18.1

[LENGTH] 132  
 [MW] 13993.36  
 [pI] 8.42  
 [HOMOL] PIR:S36081 probable carrier protein RIM2, mitochondrial - yeast (Saccharomyces cerevisiae) 7e-19  
 [FUNCAT] 07.16 purine and pyrimidine transporters [S. cerevisiae, YBR192w] 3e-20  
 [FUNCAT] 08.04 mitochondrial transport [S. cerevisiae, YBR192w] 3e-20  
 [FUNCAT] 30.16 mitochondrial organization [S. cerevisiae, YBR192w] 3e-20  
 [FUNCAT] 02.13 respiration [S. cerevisiae, YBR192w] 3e-20  
 [FUNCAT] 01.05.07 carbohydrate transport [S. cerevisiae, YPR021c] 3e-10  
 [FUNCAT] 07.07 sugar and carbohydrate transporters [S. cerevisiae, YPR021c] 3e-10  
 [FUNCAT] 07.99 other transport facilitators [S. cerevisiae, YEL006w] 1e-09  
 [FUNCAT] 01.07.10 transport of vitamins, cofactors, and prosthetic groups [S. cerevisiae, YIL006w] 3e-09  
 [FUNCAT] 07.04.07 anion transporters (cl, so4, po4, etc.) [S. cerevisiae, YKL120w] 2e-08  
 [FUNCAT] 01.03.19 nucleotide transport [S. cerevisiae, YPR011c] 3e-08  
 [FUNCAT] 04.05.03 mrna processing (splicing) [S. cerevisiae, YKR052c] 4e-08  
 [FUNCAT] 01.05.04 regulation of carbohydrate utilization [S. cerevisiae, YJR095w] 2e-07  
 [FUNCAT] 01.01.07 amino-acid transport [S. cerevisiae, YOR130c] 5e-05  
 [FUNCAT] 07.10 amino-acid transporters [S. cerevisiae, YOR130c] 5e-05  
 [FUNCAT] 01.04.07 phosphate transport [S. cerevisiae, YJR077c] 7e-05  
 [FUNCAT] 13.04 homeostasis of other ions [S. cerevisiae, YJR077c] 7e-05  
 [BLOCKS] BL00215B Mitochondrial energy transfer proteins  
 [BLOCKS] BL00215A Mitochondrial energy transfer proteins  
 [PIRKW] duplication 6e-09  
 [PIRKW] transmembrane protein 6e-09  
 [PIRKW] mitochondrial inner membrane 4e-07  
 [PIRKW] transport protein 5e-06  
 [PIRKW] mitochondrion 7e-08  
 [PIRKW] chloroplast 3e-08  
 [SUPFAM] Btl protein 3e-08  
 [SUPFAM] ADP,ATP carrier protein repeat homology 4e-09  
 [SUPFAM] Caenorhabditis probable carrier protein c2 4e-09  
 [SUPFAM] probable carrier protein YPR021c 6e-09  
 [PROSITE] LEUCINE\_ZIPPER 1  
 [PROSITE] MITOCH\_CARRIER 1  
 [PFAM] Mitochondrial carrier proteins  
 [KW] Alpha\_Beta

SEQ MSQRDTLVHLFAGGCGGTVGAILTCPLEVVKTRLQSSSVTLYISEVQLNTMAGASVNRVV





DKFZphtes3\_21d4

group: signal transduction

DKFZphtes3\_21d4 encodes a novel 464 amino acid putative GTP exchanging factor related to RCC1.

RCC1 (regulator of chromosome condensation) is a eukaryotic protein which binds to chromatin and interacts with ran, a nuclear GTP-binding protein. RCC1 promotes the exchange of bound GDP with GTP, acting as a guanine-nucleotide dissociation stimulator.

The new protein can find application in the regulation of gene expression by activation of nuclear GTP-binding proteins. The X-linked retinitis pigmentosa is a result of a defect GTPase regulator, which contains a RCC1-type repeat.

similarity to RCC1-like G exchanging factor RLG

complete cDNA, complete cds, EST hits

Sequenced by LMU

Locus: /map="20"

Insert length: 2321 bp

Poly A stretch at pos. 2293, polyadenylation signal at pos. 2262

```
1 GGGTCACGCA AGATGGCGGC GCCCAGAGGC TGCTGAGGCG CGGAACGGAG
51 GATGGCGCTG GTGGCGTTGG TGGCTGGGGC TCGGCTGGGG CGGCGGCTGA
101 GCGGGCCGGG GCTGGGGCGA GGGCACTGGA CGGCGGCCAG GCGTCCCGG
151 AGCCGGCCGG AAGCGGCAGA AGCCGAGGCG GAGGTGCCCG TGGTCCAGTA
201 CGTGGGCGAG CGCGCTGCCC GCGCCGATCG CGTCTTCGTG TGGGGCTTCA
251 GCTTCTCGGG GCGCTGGGCG GTGCCTTCCT TTGTGGTGCC CAGCTCCGGG
301 CCCGGGCCCC GCGCCGGCGC CCGACCGCGC CGCAGGATCC AGCCCGTGCC
351 CTATCGCCTG GAGCTGGACC AAAAGATTTC ATCTGCTGCT TCGCGCTATG
401 GATTACACACT GCTGTCTCTT AAGACTGCGG ATGTTACGAA AGTCTGGGGG
451 ATGGGACTCA ACAAGATTTC TCAGCTTGGG TTTCACAGGA GCCGGAAAGA
501 TAAAAACGAGG GGCTACGAGT ATGTGTTGGA GCCCTCACC GTCTCCCTGC
551 CTCTGGACAG ACCTCAGGAG ACACGGGTGC TGCAGGTCTC CTGCGGCCGA
601 GCTCACTCTC TTGTGTTGAC TGACAGGGAA GGAGTCTTCA GCATGGGAAA
651 CAAATCTTTAT GGGCAATGTG GAAGAAAGGT GGTGCGAAAT GAAATTTACA
701 GTGAAAGTCA CAGAGTCCAC AGGATGCAGG ACTTCGATGG CCAGTGGTTC
751 CAGGTGCGCT GTGGTCAGGA TCATAGTCTG TTCCTGACGG ATAAAGGAGA
801 AGTCTATTCT TGTGGATGGG GTGCTGATGG GCAAAACAGT CTGGGTCACT
851 ACAATATCAC CAGCTCGCCC ACCAAGCTGG GTGGAGACCT GCGGGGAGTG
901 AACGTTATCC AAGTTGCCAC CTACGGTGAT TGCTGCCTGG CCGTGTCCGC
951 CGACGGAGGA CTTTTTGGTT GGGGAACTC GGAGTACCTG CAGCTGGCCT
1001 CTGTCACTGA CTCACACAGG GTGAATGTGC CCCGCTGCTT ACATCTCTCA
1051 GGAGTGGGGA AGGTGCGACA GGCTGCATGC GGTGGCACGG GCTGTGCAGT
1101 GTTAAACGGA GAAGGACATG TTTTGTCTG GGGCTATGGA ATTCTTGGGA
1151 AAGGTCCAAA CCTAGTGGAA AGTGCCGTCC CTGAAATGAT TCCACCCACT
1201 CTCTTTGGCT TGACGGAGTT CAACCCAGAA ATCCAGGTTT CCCGCATCCG
1251 ATGTGGACTC AGCCACTTTG CTGCACAGAC CAACAAAGGA GAGCTGTTTG
1301 TATGGGGCAA GAACATCCGA GGGTGCCTGG GAATCGGTCT CCTGGAGGAC
1351 CAGTATTTC CATGGAGGGT GACGATGCCT GGGGAGCCTG TGGACGTGGC
1401 ATGTGGCGTG GACCACATGG TGACCTGGG CAAGTCATTC ATCTAAACCT
1451 CCCTCACCTG CTTGGGCGGC CCCGTCCCGG GAACCACTGG CACTCCTTGG
1501 CAGAGGCCAG CGCGTGCCCA GCCCCCAGGG GTTCTTGGAT GGTGGTGGCG
1551 GAGGACCTCG CGTGCACTGT GACGCTCTGT CCTGAATCCC TTAGCGGGTA
1601 CCTACCAGGA GGATCAGGGC AAGGTCCCTC TCCAGCTGCA GGTGAGGCCT
1651 GCGGAACATC GCTTGGATGG CAGCCTTTGG TGGGCCGCTG TGGCCCCGAC
1701 GTCCTCTGTT TCTCCAAGTA ACATGCGACG GTGTCTGGTG TCACGTCTCG
1751 CCTGAGAAAG CCGTCTTAGG AAAGCTTAGC TTGAACACAG TGCTCGGGAG
1801 GTTCTGCTC TGTCTGTCAT GGCAGTCTCT TGGTTTGTGT CTGGCCAAGG
1851 CCATGCGTGT GCCTCGGACC GAGCCCCAGC TTAGGCGAGG GAGTCAGGCT
1901 GGCCTCGGCC CTCGGTTTTC ATTCAAGCCA CCCTGCTCAT GGCCCTTCCT
1951 GGGCGCCTGC CACACCGCAA GCTCGCTGGG GGGACACTAG AAGCACCGTG
2001 GCCTGGGATT CCATCTGGAG CTGTCCGCGA GCACCAAGCC CAGCCTCCCA
2051 CCACGCTCAC TGCTTGCTT GGAAGAGTTA AGAAGCCCCC CAGGAAGAGA
2101 ATCAGAGCTA AGTTCTCTG CGCCGAGGGC CCCGAGCATA TCCGCCAAGG
2151 CTCAGCTGCA GTGCCAGGCG GAGGAGGAAG ATCCAGAAAT TGTGAACAAT
2201 GTTTGATTTA GTAGCGTGAC TTGCCTTTCC CTTTAAAAAC ATCTTTTACA
2251 AATCTGTCTT GGAATAAAGT CTATTTTCTG CTTTTTGGTT TTTAAAAAAA
2301 AAAAAAAAAA AAAAAAAAAA A
```

BLAST Results

-----

Entry HS203358 from database EMBL:  
human STS SHGC-31781.  
Score = 1748, P = 1.1e-72, identities = 376/394

#### Medline entries

No Medline entry

#### Peptide information for frame 1

ORF from 52 bp to 1443 bp; peptide length: 464  
Category: similarity to known protein

```

1 MALVALVAGA RLGRRLSGPG LGRGHWTAAAR RSRSRREAAE AEAEPVVOY
51 VGERAARADR VFVWGFSEFG ALGVPSFVVP SSGPGPRAGA RPRRRIQVP
101 YRLELDQKIS SAACGYGFTL LSSKTADVTK VWGMGLNKDS QLGFRHSRKD
151 KTRGYEYVLE PSPVSLPLDR PQETRVLQVS CGRAHSLVLT DREGVFSMGN
201 NSYGGQGRKV VENEIYSESH RVHRMQDFDG QVVQVACGQD HSLFLTDKGE
251 VYSCGWGADG QTGLGHYNIT SSPTKLGGDL AGVNVIVQAT YGDCCLAUSA
301 DGGFLFGWNS EYLQLASVTD STQVNVPRCL HFSGVGKVRQ AACGGTGCAV
351 LNNEGHEVFW GYGILGKGNP LVESAVPEMI PPTLFGLTEF NPEIQVSRIR
401 CGLSHFAALT NKGELFVWCK NRGCLGIGR LEDQYFPWRV TMPGEPVDVA
451 CGVDHMTLA KSFI

```

#### BLASTP hits

Entry CEW09G3\_5 from database TREMBLNEW:  
gene: "W09G3.3"; Caenorhabditis elegans cosmid W09G3  
Score = 395, P = 9.3e-37, identities = 111/330, positives = 165/330

Entry Y032\_HUMAN from database SWISSPROT:  
HYPOTHETICAL PROTEIN KIAA0032.  
Score = 309, P = 1.0e-24, identities = 96/308, positives = 143/308

Entry B38919 from database PIR:  
hypothetical protein 2 - human (fragment)  
Score = 309, P = 1.0e-24, identities = 96/308, positives = 143/308

Entry AF060219.1 from database TREMBLNEW:  
product: "RCC1-like G exchanging factor RLG"; Homo sapiens RCC1-like G  
exchanging factor RLG mRNA, complete cds.  
Score = 273, P = 4.0e-21, identities = 84/262, positives = 124/262

Entry S71752 from database PIR:  
giant protein p619 - human  
Score = 282, P = 1.1e-19, identities = 86/287, positives = 144/287

#### Alert BLASTP hits for DKFZphtes3\_21d4, frame 1

No Alert BLASTP hits found

#### Pedant information for DKFZphtes3\_21d4, frame 1

#### Report for DKFZphtes3\_21d4.1

```

{LENGTH}      464
{MW}           49997.08
{pI}           8.74
{HOMOL}        TREMBL:CEW09G3_5 gene: "W09G3.3"; Caenorhabditis elegans cosmid W09G3 5e-34

{FUNCAT}       04.07 rna transport [S. cerevisiae, YGL097w] 2e-09
{FUNCAT}       03.07 pheromone response, mating-type determination, sex-specific proteins
               [S. cerevisiae, YGL097w] 2e-09
{FUNCAT}       08.01 nuclear transport [S. cerevisiae, YGL097w] 2e-09
{FUNCAT}       04.05.05 mrna processing (5'-end, 3'-end processing and mrna degradation) [S.
               cerevisiae, YGL097w] 2e-09
{FUNCAT}       04.01.04 rrna processing [S. cerevisiae, YGL097w] 2e-09
{FUNCAT}       04.03.03 trna processing [S. cerevisiae, YGL097w] 2e-09
{FUNCAT}       30.03 organization of cytoplasm [S. cerevisiae, YGL097w] 2e-09

```

```

{FUNCAT}      30.04 organization of cytoskeleton [S. cerevisiae, YAL020c] 4e-06
{BLOCKS}      BL00870I
{BLOCKS}      BL00625B Regulator of chromosome condensation (RCC1) proteins
{BLOCKS}      BL00625A Regulator of chromosome condensation (RCC1) proteins
{PIRKW}       blocked amino end 3e-16
{PIRKW}       nucleus 3e-16
{PIRKW}       duplication 4e-08
{PIRKW}       tandem repeat 3e-16
{PIRKW}       DNA binding 3e-16
{PIRKW}       mitosis 3e-16
{PIRKW}       leucine zipper 3e-21
{SUPFAM}      pheromone response pathway component SRM1 4e-08
{SUPFAM}      WD repeat homology 3e-21
{PROSITE}     MYRISTYL 7
{PROSITE}     RCC1_2_2
{PROSITE}     AMIDATION 2
{PROSITE}     CAMP_PHOSPHO_SITE 1
{PROSITE}     CK2_PHOSPHO_SITE 5
{PROSITE}     TYR_PHOSPHO_SITE 2
{PROSITE}     GLYCOSAMINOGLYCAN 3
{PROSITE}     PKC_PHOSPHO_SITE 7
{PROSITE}     ASN_GLYCOSYLATION 2
{PFAM}        Regulator of chromosome condensation (RCC1)
{KW}          All_Beta
{KW}          LOW_COMPLEXITY 13.58 %

```

```

SEQ  MALVALVAGARLGRRLSGPGGLGRGHWTAARRSRSRREAAEAEVPPVVQYVGERAARADR
SEG  .xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx
PRD  ccchhhhhhhhhheccccccccchhhhhhhhhhhhhhhhhhhhhheeeehhhhhhhhh

```

```

SEQ  VFVWGFSGALGVPSFVVPSSGPGPRAGARPRRRIQPVPRLELDQKISSAACGYGFTL
SEG  .xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx
PRD  eeeeeccccccccccccccccccccccccccccccccccccchhhhhhhheccccccccc

```

```

SEQ  LSSKTADVTKVWGMGLNKDSQLGFHRSRKDKTRGYEYVLEPSFVSLPLDRPQETRVLQVS
SEG  .xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx
PRD  ecccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc

```

```

SEQ  CGRAHSLVLTDRGVFSMGNNYSYGQCGRKVVENEIYSESHRVHRMQDFDGGVQVQACGQD
SEG  .xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx
PRD  ccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc

```

```

SEQ  HSLFLTDKGEVYSCGWGADGQTGLGHYNITSSPTKLGGLAGVNVQVATYGDCCCLAVSA
SEG  .xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx
PRD  eeeeecccccccccccccccccccccccccccccccccccccccccccccccccccccccc

```

```

SEQ  DGGLFGWGNSEYLQLASVTDSTQVNVPRCLHFSGVGKVRQAACGGTGCAVLNGEGHVFVW
SEG  .xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx
PRD  ccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc

```

```

SEQ  GYGILGKGNPLVESAVPEMIPPTLFGLTEFNPEIQVSRIRCGLSHFAALTNKGELFVWKG
SEG  .xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx
PRD  ccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc

```

```

SEQ  NIRGCLGIGRLEDQYFPWRVTPGEPVDVACGVDMVTLAKSFI
SEG  .xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx
PRD  ccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc

```

#### Prosite for DKFZphtes3\_21d4.1

PS00001	200->204	ASN_GLYCOSYLATION	PDOC00001
PS00001	268->272	ASN_GLYCOSYLATION	PDOC00001
PS00002	17->21	GLYCOSAMINOGLYCAN	PDOC00002
PS00002	82->86	GLYCOSAMINOGLYCAN	PDOC00002
PS00002	333->337	GLYCOSAMINOGLYCAN	PDOC00002
PS00004	14->18	CAMP_PHOSPHO_SITE	PDOC00004
PS00005	34->37	PKC_PHOSPHO_SITE	PDOC00005
PS00005	122->125	PKC_PHOSPHO_SITE	PDOC00005
PS00005	147->150	PKC_PHOSPHO_SITE	PDOC00005
PS00005	190->193	PKC_PHOSPHO_SITE	PDOC00005
PS00005	219->222	PKC_PHOSPHO_SITE	PDOC00005
PS00005	246->249	PKC_PHOSPHO_SITE	PDOC00005
PS00005	410->413	PKC_PHOSPHO_SITE	PDOC00005
PS00006	34->38	CK2_PHOSPHO_SITE	PDOC00006
PS00006	147->151	CK2_PHOSPHO_SITE	PDOC00006
PS00006	190->194	CK2_PHOSPHO_SITE	PDOC00006
PS00006	290->294	CK2_PHOSPHO_SITE	PDOC00006
PS00006	317->321	CK2_PHOSPHO_SITE	PDOC00006

PS00007	209->217	TYR_PHOSPHO_SITE	PDOC00007
PS00007	208->217	TYR_PHOSPHO_SITE	PDOC00007
PS00008	9->15	MYRISTYL	PDOC00008
PS00008	20->26	MYRISTYL	PDOC00008
PS00008	133->139	MYRISTYL	PDOC00008
PS00008	238->244	MYRISTYL	PDOC00008
PS00008	277->283	MYRISTYL	PDOC00008
PS00008	302->308	MYRISTYL	PDOC00008
PS00008	344->350	MYRISTYL	PDOC00008
PS00009	12->16	AMIDATION	PDOC00009
PS00009	206->210	AMIDATION	PDOC00009
PS00626	179->190	RCC1_2	PDOC00544
PS00626	235->246	RCC1_2	PDOC00544

## Pfam for DKFZphtes3\_21d4.1

HMM_NAME	Regulator of chromosome condensation (RCC1)		
HMM	*IAAGqHHTVCLTqDGRVYtWG*		
	+A GQ+H++ LT++G VY++G		
Query	235	VACGQDHSFLFLTDKGEVYSCG	255

DKFZphtes3\_21j15

-----

group: transcription factors

DKFZphtes3\_21j15 encodes a novel 898 amino acid protein with similarity human NY-CO-33 protein.

NY-CO-33 is a protein recognised by autologous antibodies of human colon cancer patients. The novel protein contains 4 C2H2 Zinc fingers and is a new putativ transcription factor.

The new protein can find application in modulating/blocking the expression of genes controlled by this transcription factor.

strong similarity to "NY-CO-33"

complete cDNA, complete cds, potential start at bp 27, EST hits

Sequenced by LMU

Locus: unknown

Insert length: 4407 bp

Poly A stretch at pos. 4321, polyadenylation signal at pos. 4301

```

1 CGCTGCAGCA GGTGTCACAG AGCCGCATGC TCCCGGAGCC CAGCCTCTTC
51 AGCACCGTGC AGCTGTACCG GCAGAGCAGC AAGCTCTATG GCTCCATCTT
101 CACGGGGGCC AGCAAGTTCC GCTGTAAGGA CTGCAGCGCT GCCTACGACA
151 CCCTGGTGGG GTTGACAGTG CACATGAACG AGACGGGGCA TTACCGCGAC
201 GACAACCATG AGACCGATAA CAACAACCCC AAGCGCTGGT CCAAGCCTCG
251 CAAACGCTCC TTGCTGGAAA TGGAAAGGAA GGAAGACGCC CAGAAGGTGC
301 TGAAGTGCAT GTACTGTGGC CACTCCTTTG AGTCCCTGCA GGATTGTAGT
351 GTCCATATGA TCAAAACAAA ACATAACCAA AAAGTGCCTC TGAAGGAACC
401 CGTCACTCCT GTCGCCGCCA AAATCATCCC TGCCACTCGG AAGAAAGCTT
451 CCCTGGAGCT GGAGCTCCCC AGTCCCCAG ATTCCACAGG TGGAAACCCC
501 AAAGCCACCA TCTCAGACAC CAACGATGCA CTTCAGAAGA ACTCCAACCC
551 TTACATCAGC CCAATAATC GGTACGGCCA CCAGAATGGG GCCAGCTATG
601 CATGGCACTT TGAGGCCCGG AAGTCGCAGA TCCTGAAGTG CATGGAGTGT
651 GGGAGCTCGC ATGACACCCT GCAGGAGCTC ACTGCCACA TGTGGTCAC
701 TGGCCACTTC ATCAAGGTCA CCAACTCTGC TATGAAAAAG GGGAAAGCCA
751 TTGTGGAGAC GCCTGTCACA CCTACCATCA CAACCTGCTG GGATGAGAAG
801 GTCCAGTCCG TGCCCTTGGC AGCCACCACC TTCACGTCCC CCTCCAATAC
851 ACCTGGCAGC ATCTCCCCAA AACTGAATGT GGAGGTCAAG AAGGAAGTCG
901 ACAAGGAGAA AGCGGTCACT GACGAGAAAC CTAAGCAAAA AGACAAGCCT
951 GGCGAAGAAG AGGAGAAGTG TGACATCTCT TCCAAATACC ATTACTTGAC
1001 TGAAAAATGAC TTAGAAGAGA GTCCCAAGGG GGGGCTTGAT ATCCTCAAAT
1051 CCTTGGAAAA CACAGTGACA TCCGCAATCA ACAAGGCCCA GAACGGCACT
1101 CCTAGCTGGG GGGGCTATCC CAGCATCCAT GCCGCCTACC AACTTCCCAA
1151 CATGATGAAG TTGTCCCTGG GCTCGTCGGG GAAGAGCAGC CCCCTGAAAC
1201 CCATGTTTGG CAACAGTGAG ATTGTCTCCC CGACGAAAAA CCAGACCCCTG
1251 GTCTCTCCAC CCAGCAGCCA GACGTCCCCC ATGCCCAAGA CAAACTTTCA
1301 TGCCATGGAG GAGCTGGTGA AAAAGGTCAC TGAGAAAGTT GCCAAAGTGG
1351 AGGAGAAGAT GAAGGAGCCG GATGGGAAGC TTTCCCGGCC CAAGCGGGCC
1401 ACTCCCTCCC CATGTAGCAG CGAAGTCGGG GAACCCATCA AGATGGAGGC
1451 ATCCAGCGAT GGGGGCTTCC GCAGCCAGGA GAACAGCCCC AGCCCCCGGC
1501 GGGATGGGTG CAAGGATGGG AGCCCCCTCG CTGAGCCGGT GGAGAATGGC
1551 AAGGAGCTGG TGAAGCCCTT AGCCAGCAGT TTGAGTGGCA GCACGGCCAT
1601 CATCACCAGC CACCCGCTTG AACAGCCTTT TGTAAACCTT TTGAGCGCCC
1651 TGCACTCAGT CATGAACATT CACCTGGGCA AGGCCGCCAA GCCCTCCCTG
1701 CCTGCCCTGG ACCCATGAG CATGCTTTTC AAGATGAGCA ACAGCCTGGC
1751 GGAGAAGGCT GCTGTGGCCA CCCCAGCCGC CCTGCAGTCC AAGAAGGCAG
1801 ACCACCTCGA CCGCTATTTC TACCACGTCA ACAACGACCA GCCCATAGAC
1851 TTGACAAAAG GGAAGAGTGA CAAAGGCTGC TCCTTGGGTT CAGTGTCTCT
1901 GTCACCCACG TCCACAGCCC CGGCAACCTC CTCATCCACG GTGACAACGG
1951 CAAAGACATC TGCCGTCGTA TCATTCTATG CAAACTCGCC GCTACGCGAG
2001 AATGCCTTGT CAGATATATC CGATATGCTG AAGAACTTGA CAGAGAGCCA
2051 CACGTCAAAA TCCTCCACTC CTTCCAGCAT CTCCGAGAAG TCTGACATTG
2101 ACGGGGCCAC TCTGGAGGAG GCTGAGGAGT CGACGCCCGC CCAGAAAGAG
2151 AAGGGCCGCC AGTCAAACCT GAACCCCCAG CACCTCCTGA TCCTCCAGGC
2201 CCACTTTGCC GCCAGCCTCC GGCAGACCTC AGAAGGGAAG TACATCATGT
2251 CAGACCTGAG CCCCAGGAG CGGATGCATA TCTCCAGGTT CACCGGGCTG
2301 TCCATGACCA CCATCAGCCA CTGGCTGGCC AACGTGAAAT ACCAGCTTCG
2351 AAGGACAGGT GGAACAAAGT TCCTCAAAAA CTTGGACACT GGCCACCCCG
2401 TCTTCTTTTG TAACGATTGT GCGTCCCAAA TCAGGACTCC TTCCACGTAC
2451 ATCAGTCACC TAGAGTCACA CTTAGGCTTC CGGCTACGGG ACTTATCCAA
2501 ACTGTCCACC GAACAGATTA ACAGTCAGAT AGCACAAACC AAGTCACCGT
2551 CAGAAAAAAT GGTGACGTCC TCCCCGAGG AAGACCTGGG GACTTCCTAT
2601 CAGTGCAAAC TTTGCAATCG GACCTTTGCC AGCAAGCAGC CTGTTAAACT

```

```

2651 TCACCTTAGC AAAACACACG GGAATCTCC GGAAGACCAC CTTCTGTATG
2701 TCTCTGAGTT AGAGAAGCAG TAGCATTTGC TTTTGATAGA AAGGACTGCA
2751 GTTTGCTTTG AGGGAACTG TGGAAAGCAC CTTCAGGCC CCCTCTGACTT
2801 GTTGTCTTGC GCACATGTTT TTATTTTAAC TGCAGAGAAT CACTCTGGGC
2851 TGGACTGTTT TGTATACTG TACAGTGTTC AATAGAGGTG CATAATCAGC
2901 TGTTGTACTT GGTAAAAATAT GAAGGTAAAA ATGCAGTGGT AAGTGTGTTGG
2951 AACTTTGTGT AAACGGGATT TAGTTGTGAG CATCCTCCCG ATGCTTCAAG
3001 CTGCATGCAT TAACAGACAG TTTAATTAAG CATTTATAAC GGAATCAGGC
3051 ACACCTTTTC CACGAGACTC GAGTGTGCTG GCATTTCTCA CCCTTTTCATC
3101 TTTAGCCCTC TGAGTACTTT GAAGCACTTT TGCATTAATT TGGTAAAAAA
3151 ATAAAAATAA ATAATAATAA TGTATGAAGC TCTGTTTTT AAACCTCTTA
3201 CCAGCTTAGT TATAATGAAT AATATGAACC TCCATTTATG CAGGCTGCA
3251 GGGGTATAAC ACGCCTTGAA ATTTAAAGA ATATTATTTT CACATTGAAA
3301 CATAGATGTA TATATTGTAT AGATTTCAGA CTCTCTTATG AAAAAAATG
3351 TGATTGTGGT TAAATGACCT TTTCTTGCA TTTATAGCAA CAGTGTTTTA
3401 TGCACCTGCT ATGCTCTGGG CATAAGCTGT GCCTATGTAT AGTGTATATT
3451 TCTTTTTTTC TTTTTTTTAA GGTCTATGGG TTTTGTGTTT TACATGCAA
3501 CATTGTAAAT TATACAGAAG ATACCACAGA TAGCATTTAT AAAGTATACA
3551 GAAACATTAT CTGAAAGCAA AGTATGATAG TTTGTTTTC TATACAGTAC
3601 ATCTATATTG ATAGAGGTTT ATGTTTAAAT TATACATATT TATTAGCATC
3651 ATATTGTCAAT TTGTTTTGAG CAGTCTGAAT AAACGAGACC GGGAAAGACA
3701 TCCCTGGCAG GCATCAGAAC TATTTTGACAT ATGATTTTAA AAGGTATTTA
3751 TTAGAAATCA AAGAACACTC AAAATAAACT CAGTGTCTAA AGGGTTAAGT
3801 CTATTTGAAA AGGTTAAAAA AAAGAACAAA AAAAAAATAA GAACTTGATC
3851 TGTATTTTCT AAACATTGAT AAAGCCTTTA AAATGTTTGT ACTGTAATAC
3901 TTTGCTTAAA AGTCATGAGG CATTCCTGTA TCCAACCTCT TTCACTTATT
3951 TATAAGCCTT CTTGTTTGT ATTCCATATT GTAGGATGCC TTTCTATTTT
4001 AATTGGTAACT TTTCTGTTTT GTTCTTCTTA ATTATTCTCC CAAGATCCCA
4051 CACTCGACGT TTATCTTTAG GCTTATGAAA GGTAAACCCGT GGTATCCGGC
4101 TCTCCAAATG ATTCTGTTCT TCTCCATTTT TGGCAGTTAA TTGCAGAAAG
4151 TAACTGACAG CTGACACCAT ATGAGAACCT TTGTATAAAA TATTGGCATG
4201 TAAACAGCAC AGACACCGTA ACACACTCTG TGCCCTGTTT GGTGTTGAC
4251 AATGAAGCAC CATTATGTGA CTCTTCATAT AACCCCTTTT TCTACGGCAG
4301 CATTAAAAAT GTCTTTTTCG TATAAAAAAA AAAAAAATAA AAAAAAATAA
4351 AAAAAAATAA AAAAAAATAA AAAAAAATAA AAAAAAATAA AAAAAAATAA
4401 AAAAAA

```

## BLAST Results

-----

No BLAST result

## Medline entries

-----

No Medline entry

## Peptide information for frame 3

-----

ORF from 27 bp to 2720 bp; peptide length: 898

Category: strong similarity to known protein

```

1 MLPEPSLFST VQLYRQSSKL YGSIFTGASK FRCKDCSAAY DTLVELTVHM
51 NETGHYRDDN HETDNNPNKR WSKPRKRSLL EMEGKEDAQK VLKCMYCGHS
101 FESLQDLSVH MIKTKHYQKV PLKEPVPVPA AKIIPATRRK ASLELELPSS
151 PDSTGGTPKA TISDTNDALQ KNSNPYITPN NRYGHQNGAS YAWHFEARKS
201 QILKCMCEGS SHDTLQELTA HMMVTGHFIK VTNSAMKKKG PIVETPVPTPT
251 ITTLLDEKQV SVPLAATTFT SPSNTPASIS PKLNVEVKKE VDKEKAVTDE
301 KPKQKDKPGE EEEKCDISSK YHYLTENDLE ESPKGGDLIL KSLENTVTSA
351 INKAQNGTPS WGGYPSIHAA YQLPNMMKLS LGSSGKSTPL KPMFGNSEIV
401 SPTKNQTLVS PPSSQTSMP KTNFHAMEEL VKKVTEKVA VEEKMKEPDG
451 KLSPPKRATP SPCSSEVGEP IKMEASSDGG FRSQENSPSP PROGCKDGSP
501 LAEPVENGKE LVKPLASSLS GSTAITDHP PEQPFVNPLS ALQSVMIHL
551 GKAAPKSLPA LDPMSMLFKM SNSLAEKAAV ATPPLQSKK ADHLDRYFYH
601 VNNQPIDLT GKSDKGCSL GSVLLSPTST APATSSSTVT TAKTSAVVSF
651 MSNSPLRENA LSDISDMLKN LTESHTSKSS TPSSISEKSD IDGATLEEAE
701 ESTPAQKRKG RQSNWNPOHL LILQAQFAAS LRQTSEGKYI MSDLSPOERM
751 HISRETFGLSM TTISHWLANV KYQLRRGTGG KFLKNLDTGH PVFFCNDCAS
801 QIRTPSTYIS HLESHLGFERL RDLKSLSTEQ INSQIAQTKS PSEKMTSSP
851 EEDLGTSYQC KLCNRTFASK HAVKHLHLSKT HGKSPEDHLL YVSELEKQ

```

## BLASTP hits

No BLASTP hits available

## Alert BLASTP hits for DKFZphtes3\_21j15, frame 3

TREMBL:AF039698\_1 gene: "NY-CO-33"; product: "antigen NY-CO-33"; Homo sapiens antigen NY-CO-33 (NY-CO-33) mRNA, complete cds., N = 1, Score = 1039, P = 5.5e-105

PIR:A38437 probable homeotic protein tsh - fruit fly (Drosophila melanogaster), N = 3, Score = 158, P = 7.2e-09

TREMBL:CE33058\_1 gene: "unc-89"; product: "UNC-89"; Caenorhabditis elegans UNC-89 (unc-89) gene, complete cds., N = 2, Score = 175, P = 3.3e-07

>TREMBL:AF039698\_1 gene: "NY-CO-33"; product: "antigen NY-CO-33"; Homo sapiens antigen NY-CO-33 (NY-CO-33) mRNA, complete cds.  
Length = 687

## HSPs:

Score = 1039 (155.9 bits), Expect = 5.5e-105, P = 5.5e-105  
Identities = 244/504 (48%), Positives = 319/504 (63%)

Query: 170 QKNSNPYITPNNRYGHQNGASYAWHFARKSQILKCEGSSHDTLQELTAHMMVTGHFI 229  
QK +NPY+TPNNRYG+QNGASY W FEARK+QILKCEGSSHDTLQ+LTAHMMVTGHF+  
Sbjct: 14 QKAANPYVTPNNRYGQNGASYTWFARKAQILKCEGSSHDTLQQLTAHMMVTGHFL 73

Query: 230 KVTNSAMKKGKPIVETPVTPTITLLDEKVQSVPLAATTFTS-PSNT----PASISPKLN 284  
KVT SA KKGK +V PV ++EK+QS+PL TT T P+++ P S +  
Sbjct: 74 KVTTSASKKQQLVLDPV-----VEEKIQSIPLPPTTHTRLPASSIKKQPDSPAGSTT 126

Query: 285 VEVKKEVDKEKA-VTDEKPKQKDKPGEEEEKCDISSKYHYLTENDLEESPKGGDLILKSL 343  
E KKE +KEK V + K K++ + EK + S+ Y YL E DL++SPKGGDLILKSL  
Sbjct: 127 SEEKKEPEKEKPPVAGDAEKIKEESEDSEKFEFSTLYPYLREEDLDDSPKGGDLILKSL 186

Query: 344 ENTVTSAINKAQNGTPSWGYPYPSIAHAYQLPMMKLSLGSSGKSTPLKPMF-GNSEIVSP 402  
ENTV++AI+KAQNG PSWGGYPYPSIAHAYQLP +K L ++ +S ++P + G + +S  
Sbjct: 187 ENTVSTAISKAQNGAPSWGYPYPSIAHAYQLPGTVK-PLPAAVQSVQVQPSYAGGVKSLSS 245

Query: 403 TKNQTLVSPSSQTSMPKTNFAMEELVKVTEKV-AKVEEKMKEPDGKLSPPKRATPS 461  
++ L+ P S T P K+N AMEELV+KVT KV K EE+ E + K S K A S  
Sbjct: 246 AEHNALLHSPGSLTPPHKSNVSAMEELVEKVTGKVNIIKEERPEKE-KSSLAKAA--S 302

Query: 462 PCSSEVGEPKMEASSDGGFRSQENSPSPRDKGCKDGSPLAEPVENGKELVKPLASSLSG 521  
P + E + K E S + Q+ P K PL NG E +K ++  
Sbjct: 303 PIAKENKDFPKTEEVSG---KPQKKGPEAETWEAKKEGLDVHTPNGTEPLKAKVTNGCN 359

Query: 522 STAIITDHPPEQPFVNPLSALQSVNMNIHLGKAAKPSLPALDPMMLFKMSNLAEKAABA 581  
+ II DH PE F+NPLSALQS+MN HLGK +KP P+LDP++ML+K+SNS+ +K  
Sbjct: 360 NLGIIMDHSPSPFVNPLSALQSIMNTHLGKVSQVSPSLDPLAMLYKISNSMLDKPVYP 419

Query: 582 TPPPLQSKKADHLDRYFYHVNNQPIDLTGKGS DK-GCSLGSVLLSPTSTAPATSSSTVT 640  
P K+AD +DRY+Y N+DQPIDLT K+ S+ + SP + S +  
Sbjct: 420 ATPV---KQADAIDRYYYE-NSDQPIDLTGKSNKPLVSSVADSVASPLRESALMDISDMV 475

Query: 641 TAKTSAVVSFMSN-SPLRENALSDISDMLKNLTE 673  
T + S S + E + +D S + L E  
Sbjct: 476 KNLTRGLTPKSSPTSTVSEKSDADGSSFEALDE 509

Score = 865 (129.8 bits), Expect = 7.4e-95, P = 7.4e-95  
Identities = 211/434 (48%), Positives = 268/434 (61%)

Query: 447 EPDGKLSPPKRATPSPCSSEVG--EPIKMEASSDGGFRSQENSPSPRDKGCKDGSPLAE 503  
E + L P TP P S V E + + + + +E P + K SP+A+  
Sbjct: 247 EHNALLHSPGSLTPPHKSNVSAMEELVEKVTGKVNIIKEERPEKEKSSLAKAASPIAK 306

Query: 504 -----P-VE--NGKELVK-PLASSLSGSTAIITD-HPPE--QPFVNPLSALQSVNMNIHLG 551  
P E +GK K P A + D H P +P ++ + + I +  
Sbjct: 307 ENKDFPKTEEVSGKPKQKKGPEAETWEAKKEGLDVHTPNGTEPLKAKVTNGCENNLGIIMD 366

Query: 552 KAAKPSLPALDPMMLFKMSNLAEKAABATPPPLQSKKADHLDRYFYHVNN---DQPID 608  
+ +PS ++P+S L + N+ K + P L D L Y ++N D+P+  
Sbjct: 367 HSPEPSF--INPLSALQSIMNTHLGKVSQVSPSL-----DPL-AMLYKISNSMLDKPV- 417

Query: 609 LTKGKSDKGCSLGSVLLSPTSTAPATSSSTVTTAKTSAVVSFMSN-SPLRENALSDISDML 668  
K S P + + S+V ++ SPLRE+AL DISDM+  
Sbjct: 418 -YPATPVKQADAIDRYYYENSQPIDLTGKSNKPLVSSVADSVASPLRESALMDISDMV 475

Query: 669 KNLTESHTSKSSTPSSISEKSDIDGATLEEA-EESTPAQKRKGRQSNWNPOHLLILQAQF 727  
KNLT T KSSTPS++SEKSD DG++ EEA +E +P KRKGRQSNWNPOHLLILQAQF





```

SEG      x.....
PRD      cccccccccccccchhhhhhhhhccccccccchhhhhhhhhhhhhhhhhcccccc

SEG      WGGYPSIAHAYQLPNMKLSLGSSGKSTPLKPMFGNSEIVSPTKNQTLVSPSSQTSMP
SEG      .....
PRD      ccccccccccccccccccccccccccccccccccccccccccccccccccccccc

SEG      KTNFHAMEELVKKVTEKVAKEEMKEPDGKLSPPKRATPSPCSSEVGEPIKMEASSDGG
SEG      .....
PRD      ccchhhhhhhhhhhhhhhhhhhhhcccccccccccccccccccccccccccccccccc

SEG      FRSQENSPSPRDGCKDGSPLAEPVENGKELVKPLASSLSGSTAIITDHPPEQPFVNPLS
SEG      .....
PRD      ccccccccccccccccccccccccccccccccccccccccccccccccccccccc

SEG      ALQSVMIHLGKAAPSLPALDPMMLFKMSNSLAEKAAVATPPPLQSKKADHLDRYFYH
SEG      .....
PRD      chhhhhhhccccccccccccchhhhhhhhhhhhhhhcccccccccccccccccccccc

SEG      VNNDQPIDLTGKSDKCSLGSVLLSPTSTAPATSSSTVTTAKTSAVVSFMSNSPLRENA
SEG      .....
PRD      eccccceeeccccccccccccceccccccccccccceeeceeeeeeccccccccchh

SEG      LSDISMLKNLTESHTSKSSTPSSISEKSDIDGATLEAEESTPAQKRKGRQSNWNPOHL
SEG      .....
PRD      hhhhhhhhhhhccccccccccccceccccccccchhhhhhhhhccccccccchh

SEG      LILQAQFAASLRQTSEKGYIMSDLSPQERMHISRFTGLSMTTISHWLANVKYQLRRTGGT
SEG      .....
PRD      hhhhhhhhhhhccccceeeccccchhhhhhhccccchhhhhhhhhhhhhhhcccc

SEG      KFLKNLDTGHPVFVFNDCASQIRTPSTYISHLESHLGFRLRDL SKLSTEQINSQIAQTKS
SEG      .....
PRD      ceccccccccceccccceeeccccchhhhhhhhhhhhhhhhhcchhhhhhhhhhhcc

SEG      PSEKMVTSSPEEDLGTSYQCKLCNRTFASKHAVKLHLSKTHGKSPEDHLLYVSELEKQ
SEG      .....
PRD      ccceeeccccccccceehhhhhhhhhhhhhhhhhccccccccccccceeecccccc

```

## Prosites for DK72phtes3\_21j15.3

PS00001	51->55	ASN_GLYCOSYLATION	PDOC00001
PS00001	405->409	ASN_GLYCOSYLATION	PDOC00001
PS00001	670->674	ASN_GLYCOSYLATION	PDOC00001
PS00001	864->868	ASN_GLYCOSYLATION	PDOC00001
PS00004	69->73	CAMP_PHOSPHO_SITE	PDOC00004
PS00004	75->79	CAMP_PHOSPHO_SITE	PDOC00004
PS00004	139->143	CAMP_PHOSPHO_SITE	PDOC00004
PS00004	432->436	CAMP_PHOSPHO_SITE	PDOC00004
PS00004	456->460	CAMP_PHOSPHO_SITE	PDOC00004
PS00005	17->20	PKC_PHOSPHO_SITE	PDOC00005
PS00005	137->140	PKC_PHOSPHO_SITE	PDOC00005
PS00005	157->160	PKC_PHOSPHO_SITE	PDOC00005
PS00005	280->283	PKC_PHOSPHO_SITE	PDOC00005
PS00005	318->321	PKC_PHOSPHO_SITE	PDOC00005
PS00005	332->335	PKC_PHOSPHO_SITE	PDOC00005
PS00005	384->387	PKC_PHOSPHO_SITE	PDOC00005
PS00005	435->438	PKC_PHOSPHO_SITE	PDOC00005
PS00005	588->591	PKC_PHOSPHO_SITE	PDOC00005
PS00005	614->617	PKC_PHOSPHO_SITE	PDOC00005
PS00005	641->644	PKC_PHOSPHO_SITE	PDOC00005
PS00005	676->679	PKC_PHOSPHO_SITE	PDOC00005
PS00005	686->689	PKC_PHOSPHO_SITE	PDOC00005
PS00005	730->733	PKC_PHOSPHO_SITE	PDOC00005
PS00005	842->845	PKC_PHOSPHO_SITE	PDOC00005
PS00006	42->46	CK2_PHOSPHO_SITE	PDOC00006
PS00006	78->82	CK2_PHOSPHO_SITE	PDOC00006
PS00006	103->107	CK2_PHOSPHO_SITE	PDOC00006
PS00006	149->153	CK2_PHOSPHO_SITE	PDOC00006
PS00006	161->165	CK2_PHOSPHO_SITE	PDOC00006
PS00006	210->214	CK2_PHOSPHO_SITE	PDOC00006
PS00006	214->218	CK2_PHOSPHO_SITE	PDOC00006
PS00006	253->257	CK2_PHOSPHO_SITE	PDOC00006
PS00006	325->329	CK2_PHOSPHO_SITE	PDOC00006
PS00006	573->577	CK2_PHOSPHO_SITE	PDOC00006
PS00006	684->688	CK2_PHOSPHO_SITE	PDOC00006
PS00006	689->693	CK2_PHOSPHO_SITE	PDOC00006
PS00006	695->699	CK2_PHOSPHO_SITE	PDOC00006
PS00006	745->749	CK2_PHOSPHO_SITE	PDOC00006

PS00006	810->814	CK2_PHOSPHO_SITE	PDOC00006
PS00006	840->844	CK2_PHOSPHO_SITE	PDOC00006
PS00006	848->852	CK2_PHOSPHO_SITE	PDOC00006
PS00006	884->888	CK2_PHOSPHO_SITE	PDOC00006
PS00006	893->897	CK2_PHOSPHO_SITE	PDOC00006
PS00007	732->740	TYR_PHOSPHO_SITE	PDOC00007
PS00007	883->892	TYR_PHOSPHO_SITE	PDOC00007
PS00008	22->28	MYRISTYL	PDOC00008
PS00008	156->162	MYRISTYL	PDOC00008
PS00008	188->194	MYRISTYL	PDOC00008
PS00008	362->368	MYRISTYL	PDOC00008
PS00008	479->485	MYRISTYL	PDOC00008
PS00008	494->500	MYRISTYL	PDOC00008
PS00008	498->504	MYRISTYL	PDOC00008
PS00008	617->623	MYRISTYL	PDOC00008
PS00008	757->763	MYRISTYL	PDOC00008
PS00028	795->816	ZINC_FINGER_C2H2	PDOC00028
PS00028	860->882	ZINC_FINGER_C2H2	PDOC00028
PS00028	33->56	ZINC_FINGER_C2H2	PDOC00028
PS00028	94->117	ZINC_FINGER_C2H2	PDOC00028

## Pfam for DKF2phtes3\_21j15.3

HMM\_NAME Zinc finger, C2H2 type

HMM \*CpwPDCgKtFrrwsNlrRHMR..T.H\*  
 C++ C ++ + +L+ HM+ H  
 Query 33 CKD--CSAAYDTLVELTVHMMNET-GH 55

26.69 (bits) f: 94 t: 116 Target: dkfzphes3\_21j15.3 strong similarity to "NY-CO-33"  
 Alignment to HMM consensus:

Query \*CpwPDCgKtFrrwsNlrRHMR..T.H\*  
 C + CG +F + +L HM+ H  
 dkfzphes3 94 CMY--CGHSFESLQDLSVHMIKT-KH 116

Query f: 795 t: 815 Target: dkfzphes3\_21j15.3 strong similarity to "NY-CO-33"  
 Alignment to HMM consensus:

HMM \*CpwPDCgKtFrrwsNlrRHMRTH\*  
 C++ C R++S+++ H+ +H  
 Query 795 CND--CASQIRTPSTYISHLESH 815

27.12 (bits) f: 860 t: 881 Target: dkfzphes3\_21j15.3 strong similarity to "NY-CO-33"  
 Alignment to HMM consensus:

Query \*CpwPDCgKtFrrwsNlrRHMR.T.H\*  
 C+ C++TF +++ + H+ H  
 dkfzphes3 860 CKL--CNRTFASKHAVKLHLSK-TH 881

DKFZphtes3\_21116

group: intracellular transport and trafficking

DKFZphtes3\_21116 encodes a novel 66 amino acid protein nearly identical to rat ribosome attached membrane protein 4 (ramp4).

The novel protein seems to be the human orthologue of rat ramp 4. Ramp4 is involved in the regulation of translocation of proteins into endoplasmic reticulum, e.g. of the MHC class II associated invariant (gamma) chain.

The new protein can find application in modulation of protein translocation into the endoplasmic reticulum.

identical to rat ribosome attached membrane protein 4

ORF Bp 316-513 (66 aa) see BLASTX

Sequenced by LMU

Locus: unknown

Insert length: 2488 bp

Poly A stretch at pos. 2464, polyadenylation signal at pos. 2442

```
1 CTTCTCTTT CACTCCGCGC TCACGGCGGC GGCCAAAGCG GCGGCGACGG
51 CCGCGCGAGA ACGACCCGGC GGCCAGTTCT CTTCTCTCTG CGCACCTGCC
101 CCGCTCGGTC AGTCAGTCGG CGGCCGGCGC CCGGCTTGTG CTCAGACCTC
151 GCGCTTGGCG CGCCAGGCC CAGCGGCCGT AGCTAGCGTC TGGCCTGAGA
201 ACCTCGGCGC TCCGGCGGCG CGGGCACCAC GAGCCGAGCC TCGCAGCGGC
251 TCCAGAGGAG GCAGGCGAGT GAGCGAGTCC GAGGGGTGGC CGGGGCAGGT
301 GGTGGCGCCG CGAAGATGGT CGCCAAGCAA AGGATCCGTA TGGCCAACGA
351 GAAGCACAGC AAGAACATCA CCCAGCGCGG CAACGTCGCC AAGACCTCGA
401 GAAATGCCCC CGAAGAGAAG GCGTCTGTAG GACCCGTGTT ATTGGCTCTC
451 TTCATTTTGG TTGTCTGTGG TTCTGCAATT TTCCAGATTA TTCAAAGTAT
501 CAGGATGGGC ATGTGAAGTG ACTGACCTTA AGATGTTTCC ATTCTCTCTG
551 GAATTTTAAC TTGAACCTCAT TCCTGATCTT TGATACCCGT GTTGAAAAACA
601 ATTCACTAAA GCATCCTGCC TCAGAATGAC TTTCTATCA TGCTTCATGT
651 GTCATTCCAA GGTTCCTTCA TGAGTCATTC CAAGTTTCT AGTCCATACC
701 ACAGTGCCCT GCAAAAAACA CCACATGAAT AAAGCAATAA AATTTGATTG
751 TTAAGATACA GTAGTGGACC CTACTTATTC AGTCAATTAA GAGTAAGTTT
801 TTTTATGTGG TTATTAAAAC AGTATGAACA ATTAGTCTAA CTCTGCATAG
851 ACAGGTGCTA GATTTTGTTA ACCCAAATGT ATAAGTGCAG TTAGCTTAAA
901 TTACAATTTG AAGTCTTGTG GTTTTATAT AGCTAGGCAC TTTATTACTC
951 TTTTGAACCT AAAGCACACT CCTTATAGG TTCATGTAAC TGTCTGTAA
1001 TAAGGTGCTT ATAAATGGAA CAACATACACA GCCTAGTTTT GCCACAACCT
1051 TTAGCATCTA AAAAGTTTAA AAAGCTTCTA AATGTCTAAT ATAAAGGGAG
1101 ATGCTTATAG CCACAACATC TATTTTACCA ATATTGTTTC CATTACACTA
1151 CCTTGGATTG TGCATGAGTG AGTATAGTAA CCCAAGATGC CATAAAAAAA
1201 AACTTGTATG TTTTCTGACT TAATTAGTTA CTGTGGTTTC ACTAAAAGCT
1251 ACCGTGGTGG AGTGAAGTCA GTCAGGGAAG GTTTGTTTAT GTTACATTTA
1301 TTTCAACAGA ACTATTTTAA TATATCAAAG GGGTTTACTA TGCCAAACAA
1351 AATTCTAGGG AAAAATACTG CTAAAAATGG ATGCCCTATC AGAACATGCT
1401 GTTGAGTCCA ATGTGCCATA AGACATTTTA GCATGTTAAA TAGCACTTTT
1451 AATAGCAAAA AAAGGCACAT CAACTGCGAA GTTATCCTTA GTTTGCAAAAT
1501 GCTTTTCTTA GATTAATGAT TTTTCAATCA TTAGGGTACT AGACACATCA
1551 GCCTAAAGTG GCATCTGGAA TTGAATGGAT TTACTGATAA TGATCAGTCT
1601 TTAGTCTTCC CTTTGTATTA TGACTTTTATA GGTATGATT GATCAAATTT
1651 ACGTTTTTACT AATGGTAAGG GTGAGGGTCA TAGGGCAGGT TTTGGGTTTT
1701 CTAGTACTGT TGAAAACCTG AAGTATTGGC TATTTGTATA CTTAGCCATA
1751 ACTTGGTGAA AAAAACCTG AGCAGTGTCT ATGTATTAAT GCGTTGGAAA
1801 GAAAGCTGCT TGTGTTTGGT TTGTTAATTG CCTCAGGATA TTTCTTTTAA
1851 AATAAGCTGT TTTAAGAGGA ACAGAAGGGA AATCTGCTAC CTAGTCTATA
1901 CACAGCGTGA ACCTCACAGG GGGCTTCTGA TACCCTCAA CATGGAGAAC
1951 AGTAAGGGAG CAGAGTGGTT AAGGACTTTC AGGAACTTAA CTATTCTGGA
2001 ATAAGGAATG AATCAACTGA CCTTGGGCCA GCAGGTTTTT AACTAAATTG
2051 TTAATTGCGT TTCTCACCCA GTTAATCAGT CTCTGTACTT GTTTCCTTTT
2101 GTTAAACAAG TGTCTTGGTT AACTAATTCT GTTTTATGGT TGTGCTAAAT
2151 TCATAGCAGG TGCCTTATTC TTTGCTTTTA GTCAAACCAT TCCATATCAG
2201 AATTTTCCCT GGTTTACTAT AGATATTTGG CTTTAAGTTG TTGTTTGTGT
2251 TTTTAAATGT ACAAATGTTCT GATAAATTG ACTGTTAAAT TGCTATAGCT
2301 AGCAATCATT TTACATATGT AAAAAATTGC ATTCCCTTTG TATTTTCATGT
2351 GTAAATTCACC AATTAAGTGC AGTTTATATT CAGGTGGGAT TATGCATGTT
2401 TAGGTAACG AAAGCTGTGT CTTACTTGAT TTATCTTTA AAAATAAAGT
2451 TCCCTGAATA TTTGAAAAAA AAAAAAATAA AAAAAAATAA
```

## BLAST Results

Entry HSCDN13 from database EMBL:  
H.sapiens (TL5) mRNA from LNCaP cell line  
Score = 1075, P = 5.8e-41, identities = 219/221

Entry AF100470\_1 from database TREMBLNEW:  
gene: "RAMP4"; product: "ribosome attached membrane protein 4"; Rattus norvegicus ribosome attached membrane protein 4 (RAMP4) mRNA, complete cds.  
Score = 331, P = 3.9e-28, identities = 66/66, positives = 66/66, frame +1

Entry HSG19910 from database EMBL:  
human STS A002B48.  
Score = 530, P = 2.1e-17, identities = 108/109

## Medline entries

No Medline entry

## Peptide information for frame 1

ORF from 316 bp to 513 bp; peptide length: 66  
Category: strong similarity to known protein  
Classification: Intracellular transport and traffic

1 MVAQQRIRMA NEKHSKNITQ RGNVAKTSRN APEEKASVGP WLLALFIFVV  
51 CGSAIFQIIQ SIRMGM

## BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphtes3\_21116, frame 1

TREMBLNEW:RNO238236\_1 gene: "ramp4"; product: "ribosome associated membrane protein RAMP4"; Rattus norvegicus mRNA for ribosome associated membrane protein RAMP4, N = 1, Score = 331, P = 6.2e-30

TREMBL:AF100470\_1 gene: "RAMP4"; product: "ribosome attached membrane protein 4"; Rattus norvegicus ribosome attached membrane protein 4 (RAMP4) mRNA, complete cds., N = 1, Score = 331, P = 6.2e-30

>TREMBLNEW:RNO238236\_1 gene: "ramp4"; product: "ribosome associated membrane protein RAMP4"; Rattus norvegicus mRNA for ribosome associated membrane protein RAMP4  
Length = 75

## HSPs:

Score = 331 (49.7 bits), Expect = 6.2e-30, P = 6.2e-30  
Identities = 66/66 (100%), Positives = 66/66 (100%)

Query: 1 MVAQQRIRMANEKHSKNITQRGNVAKTSRNAPEEKASVGPWLLALFIFVCGSAIFQIIQ 60  
MVAQQRIRMANEKHSKNITQRGNVAKTSRNAPEEKASVGPWLLALFIFVCGSAIFQIIQ  
Sbjct: 10 MVAQQRIRMANEKHSKNITQRGNVAKTSRNAPEEKASVGPWLLALFIFVCGSAIFQIIQ 69

Query: 61 SIRMGM 66  
SIRMGM  
Sbjct: 70 SIRMGM 75

No Pedant data available

DKFZphtes3\_21n23

group: testes derived

DKFZphtes3\_15j18 encodes a novel 148 amino acid protein with strong similarity to rat 7acomp protein.

No informative BLAST results: No predictive prosite, pfam or SCOP motife.

The new protein can find application in studying the expression profile of testis-specific genes.

strong similarity to rat 7acomp protein

on genomic level encoded by AF107885

Sequenced by LMU

Locus: /map="14q24.3"

Insert length: 3122 bp

Poly A stretch at pos. 3070, polyadenylation signal at pos. 3045

```
1  GGAAAACCTC GTGGGCTCAG CCCGGGACAA AGGGCCAGGG AAGTTGGGTG
51 GTTCTGTGCT TGGTCTGTCA ATGGAGGAGA TCAAAGTTT ACGAAGGGTG
101 AAGGAGGAGA ATGATCGGCG AGGTGGATT ATTTCGCATAT TTCCTACATC
151 TGAGACATGG GAAATATATG GGTCTACCT CGAGCATAAG ACCTCAATGA
201 ACTATATGCT GGCAACACGC CTCCTCCAGG ACAGGGGAAA CCCAAGAAGA
251 AGTTTATGTA CAGGAAGAAC ACGAATGACT GCTGATGGAG CGCCAGAATT
301 GAAGATAGAG AGTCTGAATT CAAAGGCCAA GCTGCATGCT GCACTTTACG
351 AGAGGAAGCT CCTGTCTCTG GAGGTGCGAA AACGTAGACG ACGGAGTAGC
401 AGATTGAGGG CAATGAGGCC AAAATACCCA GTGATTACCC AACCAGCTGA
451 AATGAATGTT AAAACTGAGA CAGAGAGTGA AGAGGAGGAA GAAGTCGCAT
501 TAGATAATGA AGATGAAGAA CAGGAGGCTT CCCAGGAGGA GTCTGCAGGA
551 TTTCTTAGAG AAAATCAAGC CAAATATACA CCCTCATTGA CAGCTTTGGT
601 AGAAAATACA CCCAAAGAAA ATTCCATGAA AGTTCGTGAA TGGAATAATA
651 AAGGTGGACA CTGCTGCAAA CTTGAGACTC AGGAGCTAGA GCCTAAATTT
701 AACCTGATGC AGATTCTTCA AGATAATGGC AATCTTAGCA AAATGCAGGC
751 CCGAATAGCA TTCTCTGCCT ATCTCCAGCA TGTTCAAATT CGCCTGATGA
801 AAGACAGTGG CGGTCAAGCG TTCAGTGCCA GTTGGGCTGC CAAAGAGGAT
851 GAACAGATGG AGCTGGTTGT TCGTTTCTC AAGCGAGCAT CAAATAACCT
901 CCAGCATTC A CTGAGGATGG TATTACCCAG TCGACGATTG GCACTTCTGG
951 AACGCAGAAG AATCCTGGCC CACCAGCTGG GTGACTTTAT CATTGTATAC
1001 AACAAAGGAA CAGAACAAT GGCTGAAAAG AAATCAAAGA AGAAAGTTGA
1051 GGAAGAAGAG GAAGATGGGG TGAATATGGA AAACCTTCAG GAGTTCATCA
1101 GACAAGCAAG TGAGGCTGAA CTGGAGGAGG TGTGACTTT TTATACCCAA
1151 AAGAACAAGT CTGCTAGTGT CTTCTGGGG ACTCACTCTA AAATTTCTAA
1201 GAACAACAAC AATTATTCTG ATAGTGGGGC AAAAGGTGAT CACCCTGAGA
1251 CTATAATGGA AGAAGTGAAA ATAAAGCCAC CTAACAGCA ACAGACGACA
1301 GAAATTCATT CTGATAAATT ATCTCGATTT ACCACTTCAG CAGAAAAAGA
1351 GGCAAAATTA GTTTATAGCA ATTCTCTCTC TGGTCTTACT GCTACTCTGC
1401 AGAAAATTC CAACACCCAT TTGTCTATCT TTACAACCTC TGACCTCTCT
1451 CCAGGGCCTT GCCACCATTC TTCCTTATCT CAAATTCCTT CAGCTATCCC
1501 CAGCATGCCCT CACCAGCCAA CAATTTTACT GAACACAGTC TCTGCCAGTG
1551 CTCTCTCCCTG CCTACATCCC GGGGCACAGA ACATCCCAAG CCCTACTGGC
1601 CTGCCACGCT GTCGATCAGG AAGTCACACC ATTGGTCCCT TTTCTTCTCT
1651 CCAAAGTGCT GCACACATCT ATAGCCAGAA ACTGTCTCGT CCCTCTTCAG
1701 CAAAGGCAGG ATCGTGCTAT CTAACAAGC ATCATTGAGG AATAGCCAAA
1751 ACACAAAAG AGGGAGAAGA TGCTTCTTTA TATAGCAAA GGTACAACCA
1801 AAGTATGGTT ACAGCTGAAC TTCAGCGGCT AGCTGAGAAG CAGGCAGCGA
1851 GACAGTATTC TCCATCCAGC CACATCAACC TCCTCACCCA ACAGGTAACA
1901 AACCTGAATT TGGCAACTGG CATCATAAAC AGAAGCAGTG CTTAGCTCCC
1951 CCCAACCTCT CGACCCATCA TCAGTCTTAG TGGCCCGACA TGGTCTACAC
2001 AGTCAGACCC CCAAGCTCCC GAGAATCACT CCAGCTCTCC TGGAAAGCAGG
2051 AGCCTGCAGA CAGGGGGATT TGCTGGGAA GGAGAAGTAG AAAACAACGT
2101 GTACAGCCAG GCTACAGGGG TGGTCCCCCA GCACAAGTAT CACCCACAG
2151 CAGGCAGCTA TCAGCTTCAA TTTGCCCTGC AGCAACTTGA ACAACAAAAA
2201 CTTCACTCCC GGCAGCTCCT GGACCAAGT CGAGCCCGGC ACCAGGCAAT
2251 CTTTGGCAGC CAGACACTAC CTAACCTCAA TTTATGGACA ATGAATAATG
2301 GTGCAGGTTG TAGAATTICC AGTGCCACAG CTAGTGGCCA GAAGCCAACC
2351 ACTCTGCCAC AAAAAGTGGT ACCACCTCCA AGTTCTTGCG CCTCCCTGGT
2401 TCCCAAAACC CCACCAACC ACGAACAAGT GCTCAGAAGG GCAACATCCC
2451 AGAAAGCTTC CAATACCCGC TTCAGATCCT CCTTTCAAAA CTATTTGTGG
2501 TATTTCTTCC AAGCAGTCAG CTGAAGTGA GACGACAGCC TACAACAAC
2551 TACATGCATC TGAAGTGTCT CTTGTAAATG AGCTTTTTTC AGAGCCAGAA
2601 TCATACTCTC CAGGAAATAT GGAGAAAGAA ACCTGAGGAG ATTGAAGTTT
2651 GCAAGGCACA AGGGCAAAAC TCAGACTGAA TGAATTGAA AGGGTGGGGC
2701 CAAAGATGTT GTAACCTGGG AGACTTCTCT GAAGAAAGAA AACTGTTTAA
```

```

2751 GAAACACAGA CTGAAGTCA GTACTTTTCC TTAATAGCT GAGATGACCT
2801 TCITTTACCCT GGGCTTAGGT GATTCTCATC AGGGTGACCT GAGTGGAAAGT
2851 TGGTGGTAAC GACTGTTCTG TGTCAGCACC CAGGACAGTG GTGTCTGTTA
2901 AGGCTGCCAG GGATTAGCAG GGAGGAAAGC CATCAGGACT GGGTAGCCTG
2951 GTAGCACCAA ATCCCAATTA ATGTTACCTG AACATGTGGT GAGGTCAGCC
3001 GTATGATGAA AGATGTTTAA GAGATTAATG TCAGAAGAAT ATGAAAAATAA
3051 ACACCGGCTT AAAAAATGTT AAAAAAATAA AAAAAAATAA AAAAAAATAA
3101 AAAAAAATAA AAAAAAATAA AA

```

## BLAST Results

Entry AF107885 from database EMBL:  
Homo sapiens chromosome 14q24.3 clone BAC270M14 transforming growth factor-beta 3 (TGF-beta 3) gene, complete cds; and unknown genes.  
Score = 3042, P = 3.0e-219, identities = 610/612  
5 exons matching 1893-3070

## Medline entries

No Medline entry

## Peptide information for frame 2

ORF from 71 bp to 2521 bp; peptide length: 817  
Category: strong similarity to known protein

```

1 MEEIKVLRV KEENDRRGGF IRIFPTSETW EIYGSYLEHK TSMNYMLATR
51 LFQDRGNPRR SLLTGRTRMT ADGAPELKIE SLNSKAKLHA ALYERKLLSL
101 EVRKRRRRSS RLRAMRPKYP VITQPAEMNV KTESEEEEE EVALDNEDEE
151 QEASQESAG FLRENQAKYT PSLTALVENT PKENSMKVRE WNNKGGHCCK
201 LETQELEPKF NLMQILQDNG NLSKMQARIA FSAYLQHVQI RLMKDSGGQT
251 FSASWAAAKD EQMELVVRFL KRASNQLQHS LRMVLPSSRL ALLERRRILA
301 HQLGDFIIVY NKETEQMAEK KSKKKVEEEE EDGVMNMFQ EFIRQASEAE
351 LEEVLTFYTO KNKSASVFLG THSKISKNNN NYSDSGAKGD HPETIMEEVK
401 IKPPKQQQTT EIHSDKLSRF TTSAEKEAKL VYNSSSSGPT ATLQKIPNTH
451 LSSVTTSDLS PGPCHHSSLS QIPSAIPSMF HQPTILLNTV SASASPCLHP
501 GAQNIPSPPTG LPRCRSGSHT IGFSSSFQSA AHIYSQKLSR PSSAKAGSCY
551 LNKHHSGLAK TQKEGEDASL YSKRYNQSMV TAEQLRLAEK QAARQYSPSS
601 HINLLTQQVT NLNLATGIIN RSSASAPPTL RPIISPSGPT WSTQSDPQAP
651 ENHSSSPGSR SLQTGGFAWE GEVENNVYSQ ATGVVPQHKY HPTAGSYQLQ
701 FALQOLEQQK LQSRQLLDQS RARHQAFIFS QTLFNSNLWT MNNGAGCRIS
751 SATASGQKPT TLPQKVPPPP SSCASLVPKP PPNHEQVLRN ATSQKASNTR
801 FRSSFQNYLW YFFQAVS

```

## BLASTP hits

No BLASTP hits available

## Alert BLASTP hits for DKFZphtes3\_2ln23, frame 2

TREMBL:AF064856\_1 product: "7aomp protein"; Rattus sp. 7aomp protein mRNA, complete cds., N = 1, Score = 1845, P = 2.2e-190

TREMBL:AF107885\_3 product: "unknown"; Homo sapiens chromosome 14q24.3 clone BAC270M14 transforming growth factor-beta 3 (TGF-beta 3) gene, complete cds; and unknown genes., N = 1, Score = 443, P = 5.3e-41

TREMBL:AF107885\_4 product: "unknown"; Homo sapiens chromosome 14q24.3 clone BAC270M14 transforming growth factor-beta 3 (TGF-beta 3) gene, complete cds; and unknown genes., N = 1, Score = 265, P = 8.2e-22

>TREMBL:AF064856\_1 product: "7aomp protein"; Rattus sp. 7aomp protein mRNA, complete cds.  
Length = 436

## HSPs:

Score = 1845 (276.8 bits), Expect = 2.2e-190, P = 2.2e-190  
Identities = 369/435 (84%), Positives = 395/435 (90%)

Query:	115	MRPKYPVITQPAEMNVKTTETESEEEEVALDNEDEEQEASQESAGFLRENQAKYTPSLT	174
		MRPKYPVIT PAEMN+KTETESEEEEEV LDNEDEEQEASQESAG L ENQAKYTPSLT	
Sbjct:	1	MRPKYPVITLPAEMNKTETESEEEEEVGLDNEDEEQEASQESAGSLAENQAKYTPSLT	60
Query:	175	ALVENTPKENSMKRWNNKGGHCCCKLEQLEQPFKNLMQILQDNGNLSKMQARIAFSAY	234
		+VEN+P+EN+MKV EW NKG CCK+EQE E KFNLMQILQDNGNLSK+QAR+AFSAY	
Sbjct:	61	VIVENSPRENAMKVAEWTNKGESCCKTETQEPESKFNLMQILQDNGNLSKVQARLAFSAY	120
Query:	235	LQHVQIRLMDKDSGGQTFASWAAKEDQEMELVVRFLKRASNQLQHSRLMVLPSRRLALLE	294
		LQHVQ+RL KDSGGQT S WAAKEDQEMELVVRFLKRAS+NLQHSRLMVLPSRRLALLE	
Sbjct:	121	LQHVQRLRLKDSGGQTLSPSWAAKEDQEMELVVRFLKRASNQLQHSRLMVLPSRRLALLE	180
Query:	295	RRRILAHQLGDFIIVYNKETEQMAEKSKKKVEEEEEDGVNMFQEFIRQASEAELEEV	354
		RRRILAHQLGDFI+VYNKETEQMAEKSKKK+EEEEEDGVN E+QFEFIRQASEAELEEV	
Sbjct:	181	RRRILAHQLGDFIIVVYNKETEQMAEKSKKKLEEEEEDGVNAESFQEFIRQASEAELEEV	240
Query:	355	LTFYQKNKSASVFLGTHSKISKNNNNYSDSGAKGDHPETIMEEVKIKPKKQQOTTEIHS	414
		LTFYQKNKSASVFLGTHSK SKN++YSDSGAKGDHPETI +EVKIK PKQQO TEIHS	
Sbjct:	241	LTFYQKNKSASVFLGTHSKSSKNSSSYSDSGAKGDHPETI -QEVKIKQPKQQQATEIHS	299
Query:	415	DKLSRFTTSAEKEAKLVYSNSSS--GPTATL-QKIPNTHLSSV-TTSDLSPGPGCHSSLS	470
		DKLSRFTTSA KEAKLVY+N SS GP A L Q++P+THLSS+ TTS LS GP HHSSLS	
Sbjct:	300	DKLSRFTTSAGKEAKLVYNTCCSFGPAAVLLQRLPSTHLSSII+TTTSLSSGPGHHSSLS	359
Query:	471	QISPAIPSMPHQPTILLNTVSASPLCHPGAQNI+SPPTGLPCRSGSHITGPFSSFQSA	530
		QI AIPSMPHQ +LLN V SASP +HPG N+ SP GLPCRSGS+TIGPFSSFQSA	
Sbjct:	360	QISPAIPSMPHQSALLNPVPDSASPVPVHFGTPNV-SPAGLPCRSGSYTIGPFSSFQSA	418
Query:	531	AHIYSQKLSRPSSAKAG	547
		AHIYSQKLSRPSSAKAG	
Sbjct:	419	AHIYSQKLSRPSSAKAG	435

Pedant information for DKFZphtes3\_21n23, frame 2

## Report for DKFZphtes3 21n23.2

```
[LENGTH]          817
[MW]              91522.09
[pI]              9.32
[HOMOL]           TREMBL:AF064856_1 product: "7acomp protein"; Rattus sp. 7acomp protein mRNA,
complete cds. 1e-166
[PROSITE]         MYRISTYL          6
[PROSITE]         CAMP_PHOSPHO_SITE      4
[PROSITE]         CK2_PHOSPHO_SITE      12
[PROSITE]         TYR_PHOSPHO_SITE       1
[PROSITE]         PKC_PHOSPHO_SITE      15
[PROSITE]         ASN_GLYCOSYLATION      7
[KW]              Alpha_Beta
[KW]              LOW_COMPLEXITY      13.83 %
```

[illegible]

```

SEQ      TTSAEKEAKLVYSSSSGPTATLQKIPNTHLSSVTTSDLSPGPCHHSSLQIPSAIPSPM
SEG      .....
PRD      hhhhhhhheeeccccccccceeeccccccccccccccccccccccccccccccccccccc

SEQ      HQPTILLNTVSASAPCLHPGAQNIPTGLPRCRSGSHTIGPFSSFQSAAHYISQKLSR
SEG      .....
PRD      cccceeeccccccccccccccccccccccccccccccccccccccccchhhhhhhhhccc

SEQ      PSSAKAGSCYLNKHHSGIAKTQKEGEDASLYSKRYNQSMVTAELQRLAEKQARQYSPSS
SEG      .....
PRD      cccccceeeccccccccccccccccceeeccchhhhhhhhhhhhhhhhhhhhhccc

SEQ      HINLLTQQVTNLNLATGIINRSSASAPPTLRPIISPSGPTWSTQSDPQAPENHSSSPGSR
SEG      ..xxxxxxxxxxxxx.....
PRD      cccccccccccccccccccccccccceeeccccccccccccccccccccccccccccc

SEQ      SLQTGGFAWEGEVENNVYSQATGVVPQHKYHPTAGSYQLQALQOLEQOKLQSRQLDQS
SEG      .....
PRD      cccccceeeccccccccccccccccccccccchhhhhhhhhhhhhhhhhhhhhhhhhhhhh

SEQ      RARHQAIFGSQTLPNSNLWTMNGAGCRISSATASGQKPTTLPQKVVPSPSSCASLVKPK
SEG      .....
PRD      hhhhhhhhhccccccccceeeccccceeeccccccccccccceeeccccceeecccccc

SEQ      PPNHEQVLRRATSQKASNTRFRSSFQNYLWYFFQAVS
SEG      .....
PRD      cccccchhhhhhhhhccccccccccccceeecccccc

```

## Prosites for DKFZphtes3\_21n23.2

PS00001	221->225	ASN_GLYCOSYLATION	PDOC00001
PS00001	362->366	ASN_GLYCOSYLATION	PDOC00001
PS00001	381->385	ASN_GLYCOSYLATION	PDOC00001
PS00001	434->438	ASN_GLYCOSYLATION	PDOC00001
PS00001	576->580	ASN_GLYCOSYLATION	PDOC00001
PS00001	620->624	ASN_GLYCOSYLATION	PDOC00001
PS00001	652->656	ASN_GLYCOSYLATION	PDOC00001
PS00004	106->110	CAMP_PHOSPHO_SITE	PDOC00004
PS00004	107->111	CAMP_PHOSPHO_SITE	PDOC00004
PS00004	271->275	CAMP_PHOSPHO_SITE	PDOC00004
PS00004	789->793	CAMP_PHOSPHO_SITE	PDOC00004
PS00005	64->67	PKC_PHOSPHO_SITE	PDOC00005
PS00005	109->112	PKC_PHOSPHO_SITE	PDOC00005
PS00005	180->183	PKC_PHOSPHO_SITE	PDOC00005
PS00005	185->188	PKC_PHOSPHO_SITE	PDOC00005
PS00005	280->283	PKC_PHOSPHO_SITE	PDOC00005
PS00005	287->290	PKC_PHOSPHO_SITE	PDOC00005
PS00005	322->325	PKC_PHOSPHO_SITE	PDOC00005
PS00005	359->362	PKC_PHOSPHO_SITE	PDOC00005
PS00005	414->417	PKC_PHOSPHO_SITE	PDOC00005
PS00005	535->538	PKC_PHOSPHO_SITE	PDOC00005
PS00005	543->546	PKC_PHOSPHO_SITE	PDOC00005
PS00005	561->564	PKC_PHOSPHO_SITE	PDOC00005
PS00005	572->575	PKC_PHOSPHO_SITE	PDOC00005
PS00005	629->632	PKC_PHOSPHO_SITE	PDOC00005
PS00005	793->796	PKC_PHOSPHO_SITE	PDOC00005
PS00006	35->39	CK2_PHOSPHO_SITE	PDOC00006
PS00006	132->136	CK2_PHOSPHO_SITE	PDOC00006
PS00006	134->138	CK2_PHOSPHO_SITE	PDOC00006
PS00006	136->140	CK2_PHOSPHO_SITE	PDOC00006
PS00006	154->158	CK2_PHOSPHO_SITE	PDOC00006
PS00006	180->184	CK2_PHOSPHO_SITE	PDOC00006
PS00006	347->351	CK2_PHOSPHO_SITE	PDOC00006
PS00006	394->398	CK2_PHOSPHO_SITE	PDOC00006
PS00006	422->426	CK2_PHOSPHO_SITE	PDOC00006
PS00006	455->459	CK2_PHOSPHO_SITE	PDOC00006
PS00006	561->565	CK2_PHOSPHO_SITE	PDOC00006
PS00006	643->647	CK2_PHOSPHO_SITE	PDOC00006
PS00007	563->572	TYR_PHOSPHO_SITE	PDOC00007
PS00008	195->201	MYRISTYL	PDOC00008
PS00008	248->254	MYRISTYL	PDOC00008
PS00008	510->516	MYRISTYL	PDOC00008
PS00008	557->563	MYRISTYL	PDOC00008
PS00008	746->752	MYRISTYL	PDOC00008
PS00008	756->762	MYRISTYL	PDOC00008

(No Pfam data available for DKFZphtes3\_21n23.2)



DKFZphtes3\_22c23

group: testes derived

DKFZphtes3\_22c23 encodes a novel 223 amino acid protein without similarity to known proteins.

No informative BLAST results; No predictive prosite, pfam or SCOP motif.

The new protein can find application in studying the expression profile of testis-specific genes.

unknown

complete cDNA, complete cds, 3 EST hits (two from a testis library)

Sequenced by LMU

Locus: /map="9q34"

Insert length: 1113 bp

Poly A stretch at pos. 1073, polyadenylation signal at pos. 1055

```

1 GGTGGGCAAA GGCATCTTCC TCTGGGAAGG ACTGGCACAA GCACCTGGTC
51 CTTGGGTTGT GTGCCTGGGA GGCCGGGATC AGGGCTGGCC CTCTTTCTCC
101 CTGGCAAAGC AAAACCTCCC TTTTACTACT ATCAAGGGGA AGTAACTTGA
151 AGGTGCCTGT GGCAGGCAGC ACCTTGAGCC AACAGGAACC ATTGACATGC
201 GAGGCCCAGG GCAGGCAGAC TGTGCAGTGG CCATTGGGCG GCCCCTCGGG
251 GAGGTGGTGA CCCTCCGCGT CTTGAGAGT TCTCTCAACT GCAGTGCGGG
301 GGACATGTTG CTGCTTTGGG GCCGGCTCAC CTGGAGGAAG ATGTGCAGGA
351 AGCTGTTGGA CATGACTTTC AGCTCCAAGA CCAACACGCT GGTGGTGAGG
401 CAGCGCTGCG GCGGCCAGG AGTGGGGTG CTGCTGCGGT ATGGGAGCCA
451 GCTTGCTCCT GAAACCTTCT ACAGAGAATG TGACATGCAG CTCTTTGGGC
501 CCTGGGGTGA AATCGTGAGC CCCTCGCTGA GTCCAGCCAC GAGTAATGCA
551 GGGGGCTGCC GGCTCTTCAT TAATGTGGCT CCGCACGCAC GGATTGCCAT
601 CCATGCCCTG GCCACCAACA TGGGCGCTGG GACCGAGGGA GCCAATGCCA
651 GCTACATCTT GATCCGGGAC ACCACAGCT TGAGGACCAC AGCGTTCCAT
701 GGGCAGCAGG TGCTCTACTG GGAGTCAGAG AGCAGCCAGG CTGAGATGGA
751 GTTCAGCGAG GGCTTCCTGA AGGCTCAGGC CAGCCTGCGG GGCCAGTACT
801 GGACCCTCCA ATCATGGGTA CCGGAGATGC AGGACCCTCA GTCCTGGAAG
851 GGAAGGAAG GAACCTGAGG GTCATTGAAC ATTTGTTCCG TGCTTGCCA
901 GCCCTGGAGG GTTGACCCCT GGTCTCAGT CTTTCCAATT CGAAGCTTTT
951 CCAATCTTAG GTATCTACTT TAGAGTCTTC TCCAATGTCC AAAAGGCTAG
1001 GGGGTTGGAG GTGGGGACTC TGGAAAAGCA GCCCCCATTT CCTCGGGTAC
1051 CAATAAATAA AACATGCAGG CTGAAAAAAA AAAAAAAAAA AAAAAAAAAA
1101 AAAAAAAAAA AAA

```

## BLAST Results

Entry HSAC1644 from database EMBL:  
 Genomic sequence from Human 9q34, complete sequence.  
 Score = 2072, P = 8.8e-225, identities = 422/430  
 5 exons Bp 41969-38232

## Medline entries

No Medline entry

## Peptide information for frame 2

ORF from 197 bp to 865 bp; peptide length: 223  
 Category: putative protein

```

1 MRGPGQADCA VAIGRPLGEV VTLRVLESSL NCSAGDMLLL WGRLTWRKMC
51 RKLLDMTFSS KTNLTIVVRQR CGRPGGGVLL RYGSQAPET FYRECDMQLF
101 GPWGEIVSPS LSPATSNAGG CRLFINVAPH ARIAIHALAT NMGAGTEGAN
151 ASYILIRDTL SLRTTAFHQO QVLYWESESS QAEMEFSEGF LKAQASLRGQ
201 YWTLQSWVPE MQDPQSWKKG EGT

```

## BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphtes3\_22c23, frame 2

No Alert BLASTP hits found

Pedant information for DKFZphtes3\_22c23, frame 2

## Report for DKFZphtes3\_22c23.2

```

[LENGTH]      223
[MW]           24546.19
[pI]           8.57
[PROSITE]      MYRISTYL      4
[PROSITE]      CK2_PHOSPHO_SITE      2
[PROSITE]      PKC_PHOSPHO_SITE      6
[PROSITE]      ASN_GLYCOSYLATION      2
[KW]           Alpha_Beta

SEQ    MRGPGQADCAVAIGRPLGEVVTLRVLESSLNCSAGDMLLLWGRLTWRKMCRLDLMTFSS
PRD    ccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc

SEQ    KTNTLVVRQRCGRPGGVLLRYGSQLAPETFYRECDMQLFGPWGEIVSPSLSPATSNAGG
PRD    ccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc

SEQ    CRLFINVAPHARIAIHALATNMGAGTEGANASYILIRDTHSLRTTAFHGQQVLYWESESS
PRD    ccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc

SEQ    QAEMEFSEGFLKAQASLRGQYWTLSWVPEMQDPQSWKGKEGT
PRD    hhhhhhhcchhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh

```

## Prosites for DKFZphtes3\_22c23.2

```

PS00001      31->35    ASN_GLYCOSYLATION      PDOC00001
PS00001      150->154  ASN_GLYCOSYLATION      PDOC00001
PS00005      22->25    PKC_PHOSPHO_SITE      PDOC00005
PS00005      45->48    PKC_PHOSPHO_SITE      PDOC00005
PS00005      59->62    PKC_PHOSPHO_SITE      PDOC00005
PS00005      161->164  PKC_PHOSPHO_SITE      PDOC00005
PS00005      196->199  PKC_PHOSPHO_SITE      PDOC00005
PS00005      216->219  PKC_PHOSPHO_SITE      PDOC00005
PS00006      33->37    CK2_PHOSPHO_SITE      PDOC00006
PS00006      180->184  CK2_PHOSPHO_SITE      PDOC00006
PS00008      5->11     MYRISTYL              PDOC00008
PS00008      145->151  MYRISTYL              PDOC00008
PS00008      148->154  MYRISTYL              PDOC00008
PS00008      199->205  MYRISTYL              PDOC00008

```

(No Pfam data available for DKFZphtes3\_22c23.2)

DKFZphtes3\_22g2

group: nucleic acid management

DKFZphtes3\_22g2 encodes a novel 1230 amino acid protein with nearly identical to rat TIP120.

TATA-binding protein TBP is a central component for transcriptional regulation and is a target for various transcription regulators. TBP-interacting protein 120 (TIP120) is a protein interacting with the TATA-binding protein (TBP). The novel protein is the human ortholog of rat TIP120. The novel TBP-binding protein is considered to participate in transcription regulation through the interaction with TBP.

The new protein can find application in modulation of gene transcription.

KIAA0829, complete cds, nearly identical to rat TIP120

complete cDNA, complete cds, EST hits,

Sequenced by LMU

Locus: /map="387.3 cR from top of Chr12 linkage group"

Insert length: 5387 bp

Poly A stretch at pos. 5352, polyadenylation signal at pos. 5335

```
1 GGGAGCGAGT GCGGAGCGAG TGGGAGCGAG ACGGCCCTGA GTGGAAGTGT
51 CTGGCTCCCC GTAGAGGCCCT TTCTGTACGC CCCGCCGCC ATGAGCTCGT
101 TCTCAGCGCA ACAGCGCCGT CGTTAGGCTG GCTCTGTAGC CTCGGCTTAC
151 CCCGGGACAG GCCACGCCT CGCCAGGGAG GGGGCAGCCC GTCGAGGCGC
201 CTCCCTAGTC AGCGTCGGCG TCGCGCTGCG ACCCTGGAAG CGGGAGCCGC
251 CGCGAGCGAG AGGAGGAGCT CCAGTGGCGG CGGCGGCGGC GGCAGCGGCA
301 GCGGGCAGAG GCTCCAGCAG CGCCAGCAGG CGGGATCGAG GCCGTCAACA
351 TGGGAGCGCG CTCGTACCAC ATTTCCAATT TGCTGGAAAA AATGACATCC
401 AGCGACAAGG ACTTTAGGTT TATGGCTACA AATGATTGA TGACGGAAC
451 GCAGAAAGAT TCCATCAAGT TGGATGATGA TAGTGAAAGG AAAGTAGTGA
501 AAATGATTTT GAAGTTATTG GAAGATAAAA ATGGAGAGGT ACAGAAATTTA
551 GCTGTCAAAT GTCTTGCTCC TTAGTGAGT AAAGTGAAAG AATACCAAGT
601 AGAGACAATT GTAGATACCC TCTGCACTAA CATGCTTTCT GATAAAGAAC
651 AACTTCGAGA CATTTCAGT ATTGGTCTTA AAACAGTAAT TGGAGAAC
701 CCTCCAGCTT CCAGTGGCTC TGCATTAGCT GCTAATGTAT GTAAAAAGAT
751 TACTGGACGT CTTACAAGTG CAATAGCAAA ACAGGAAGAT GTCTCTGTTC
801 AGCTAGAGAG CTTGGATATT ATGGCTGATA TGTTGAGCAG GCAAGGAGGA
851 CTTCTTGTTA ATTTCCATCC TTCAATTCTG ACCTGTCTAC TTCCCCAGTT
901 GACCAGCCCT AGACTTGCAG TGAGGAAAAG AACCATTATC GCTCTGGCC
951 ATCTGGTTAT GAGCTGTGGA AATATAGTTT TTGTAGATCT TATTGAACAT
1001 CTGTTGTCAG AGTTGTCCAA AAATGATTCT ATGTCAACAA CAAGAACCTA
1051 CATACAATGT ATTGCTGCTA TTAGTAGGCA AGCTGGTCAT AGAATAGGTG
1101 AATACCTTGA GAAGATAATT CCTTTGGTGG TAAAAATTTG CAATGTAGAT
1151 GATGATGAAT TAAGAGAGTA CTGTATTCAA GCCTTTGAAT CATTGTGAAG
1201 AAGATGTCCT AAGGAAGTAT ATCCTCATGT TTCTACCATT ATAAATATTT
1251 GTCTTAAATA TCTTACCTAT GATCCAAATT ATAATTACGA TGATGAAAGT
1301 GAAGATGAAA ATGCAATGGA TGCTGATGGT GGTGATGATG ATGATCAAGG
1351 GAGTGATGAT GAATACAGTG ATGATGATGA CATGAGTTGG AAAGTGAGAC
1401 GTGCAGCTGC GAAGTGCTTG GATGCTGTAG TTAGCACAAG GCATGAAATG
1451 CTTCCAGAAT TCTACAAGAC CGTCTCTCCT GCCTAATAT CCAGATTTAA
1501 AGAGCGTGAA GAGAAATGTA AGGCAGATGT TTTTCACGCA TACCTTTCTC
1551 TTTTGAAGCA AACTCGTCTT GTACAAAGTT GGCTATGTGA CCCTGATGCA
1601 ATGGAGCAGG GAGAAACACC TTAACAATG CTTCAGAGTC AGGTTCCCAA
1651 CATTGTTAAA GCTCTTCACA AACAGATGAA AGAAAAAGT GTGAAGACCC
1701 GACAGTGTTG TTTTAACATG TTAAGTGAAG TGTTAAATGT ATTACCTGGG
1751 GCCCTAATCT AACACATTCC TGTACTTGTA CCAGGAATCA TTTTCTCACT
1801 GAATGATAAA TCAAGCTCAT CGAATTTGAA GATCGATGCT TTGTCATGTC
1851 TATACGTAAT CCTCTGTAAC CATTCTCCTC AAGTCTTCCA TCCTCACGTT
1901 CAGGCTTTGG TTCTCCAGT GGTGGCTTGT GTTGAGAGCC CATTTTACAA
1951 AATTACATCT GAAGCACTTC TTGTTACTCA ACAGCTTGTC AAAGTAATTC
2001 GTCTTTTACA TCAGCCTTCC TCGTTTGATG CAACTCCTTA TATCAAAGAT
2051 CTATTTTACC GTACCATTA GAGATTAAAA GCAGCTGACA TTGATCAGGA
2101 AGTCAAGGAA AGGGCTATTT CCTGTATGGG ACAAAATTAT TGCAACCTTG
2151 GAGACAATTT GGGTTCTGAC TTGCCTAATA CACTTCAGAT TTTCTTGGAG
2201 AGACTAAAGA ATGAAATTAC CAGGTAACT ACAGTAAAGG CATTGACACT
2251 GATTGCTGGG TCACCTTTGA AGATAGATT GAGGCTGTT CTGGGAGAAG
2301 GGGTTCTTAT CTTGCTTCA TTTCTTAGAA AAAACCAGAG AGCTTTGAAA
2351 CTGGGTACTC TTTCTGCCCT TGATATTCTA ATAAAAAAT ATAGTGACAG
2401 CTTGACAGCT GCCATGATTG ATGCAGTTCT AGATGAGCTC CCACCTCTTA
2451 TCAGCGAAAG TGATATGCAT GTTTCACAAA TGGCCATCAG TTTTCTTACC
2501 ACTTTGGCAA AAGTATATCC CTCCTCCCTT TCAAAGATAA GTGGATCCAT
2551 TCTCAATGAA CTTATTGGAC TTGTGAGATC ACCCTTATTG CAGGGGGGAG
```

```

2601 CTCTTAGTGC CATGCTAGAC TTTTCCAAG CTCTGGTTGT CACTGGAACA
2651 AATAATTAGG GATACATGGA TTTGTTGCGC ATGCTGACTG GTCCAGTTTA
2701 CTCTCAGAGC ACAGCTCTTA CTCATAAGCA GTCTTATTAT TCCATTGCCA
2751 AATGTGTAGC TGCCCTTACT CGAGCATGCC CTAAGAGGGG ACCAGCTGTA
2801 GTAGGTCAGT TTATTCAAGA TGTCAAGAAC TCAAGGTCTA CAGATTCCAT
2851 TCGTCTCTTA GCTCTACTTT CTCTGGAGA AGTTGGGCAT CATATTGACT
2901 TAAGTGGACA GTTGGAACTA AAATCTGTAA TACTAGAAGC TTTCTCATCT
2951 CCTAGTGAAG AAGTCAAATC AGCTGCATCC TATGCATTAG GCAGCATTAG
3001 TGTGGGCAAC CTTCTGAAT ATCTGCCGTT TGTCTGCAA GAAATAACTA
3051 GTCAACCCAA AAGGCAGTAT CTTTACTTC ATTCTTGAA GGAATTATT
3101 AGCTCTGCAT CAGTGGTGGG CCTTAAACCA TATGTTGAAA ACATCTGGGC
3151 CTTATTACTA AAGCACTGTG AGTGTGCAGA GGAAGGAACC AGAAATGTTG
3201 TTGCTGAATG TCTAGGAAAA CTCACCTCTA TTGATCCAGA AACTCTCCTT
3251 CCACGGCTTA AGGGGTACTT GATATCAGGC TCATCATATG CCCGAAGCTC
3301 AGTGGTTACG GCTGTGAAAT TTACAATTTC TGACCATCCA CAACCTATTG
3351 ATCCACTGTT AAAGAACTGC ATAGGTGATT TCCTAAAAAC TTTGGAAGAC
3401 CCAGATTGTA ATGTGAGAAG AGTAGCCTTG GTCACATTTA ATTCCAGCAGC
3451 ACATAACAAG CCATCATTA TAAGGGATCT ATTGGATACT GTTCTTCCAC
3501 ATCTTTACAA TGAACAAAA GTTAGAAAGG AGCTTATAAG AGAGGTAGAA
3551 ATGGGTCCAT TTAACATAC GGTGATGAT GGTCTGGATA TTAGAAAGGC
3601 AGCATTGAGG TGTATGTACA CACTTCTAGA CAGTTGTCTT GATAGACTTG
3651 ATATCTTTGA ATTTCTAAAT CATGTTGAAG ATGTTTGAA GGACCATTAT
3701 GATATTAAGA TGCTGACATT TTTAATGTTG GTGAGACTGT CTACCCCTTG
3751 TCCAAAGTGA GACTGCAGA GGTGGACCG ACTTGTGAG CCATTACGTG
3801 CAACATGTAC AACTAAGGTA AAGGCAAACT CAGTAAAGCA GGAGTTTGAA
3851 AAACAAGATG AATTAAGCG ATCTGCCATG AGAGCAGTAG CAGCACTGCT
3901 AACCATTCCA GAAGCAGAGA AGAGTCCACT GATGAGTGAA TTCCAGTCAC
3951 ACATCAGTTC TAACCTGAG CTGGCGGCTA TCTTTGAAAG TATCCAGAAA
4001 GATTTCATCT CACTAACTT GGAATCAATG GACACTAGTT AGATGTTTGT
4051 TCACCATGGG GACCATTACA TATGACCATA CAATGCACTG AATTGACAGG
4101 TTAATCATAA GACATGGAAA GAGAAGTGTC TAAAGCTTC AAAATGTTCC
4151 ACTTTTCTTT CCTTCATGGA GACTGTTTGT TTGGCTTTCT TCCATTGTTG
4201 TTTTGTAGC ATTTATTTC AATATGTTG TTTCCATAAT CCAGAGGTTG
4251 TAAACCACT AGTGTTTTAG TGGTTACAGC AACATTTGAA ATGGAACTA
4301 AAAGTTAGGA TTTTATGGAG TATGGAGATA GGGTCCAGTA TCTATTTACC
4351 CTGTAATGTT TAGGATTAAT ATGTTAAAT TTTGTGACCA TGAATTTCTT
4401 TCTTTTATAA ATTTCTCAT TTAATAATCA AAAATCTTGC AAAACAAAAA
4451 CCATGTTTCT TTTTCTTGT TAACTTTGT TTTTCAGCAA CATAAATTGA
4501 TTTTATAGCT GCAGACAAGA ATATCCATAT AAGATTGTTT AACCATTTCA
4551 GAGAGTTTGG CAATTTTAA AAGATAATAA GGTATCATTT TTAAGTATGA
4601 AAATTAAACA TATCCCTGTT GCGCACACTA ATTTGCAATG AGTAAGTTTA
4651 CAAATATGTA TCGTCTGTAA AGCAGCATGT GCAGATTATT CATAATATAG
4701 AAGTTAAAT AAGTATTAGT GCAATTTTCA GATATTATT TTTGCACAGA
4751 AAACACATTA TCTGGAGAGA AAGAAAGGAG AATTTTGTAG ACTTGGGTTT
4801 TCTTAATGCC AGTGTGAATT TGCAGATGTT TTCAGAAAAT CAAGTCACAG
4851 TAACAATTTG CCACTTTTTT CTATTATAAA TCTTCTACT TAAATTTTGA
4901 ATATTTAGTT TTTCTCAGTT ACCCATTTGT GTGTGTGTGA TTCCACTTAG
4951 AAATTTCTAA AACCAAGATT TTCTTTCATT CCGTTTGGAT GTCTACATTC
5001 CTTATCAAAG GATATAAATA CTGTGTATGC TTTTGAATTT TATTTTATAG
5051 AAAATTCTGA AGCCAGCTAT CACAGGTTTG TTAGCTAATA ATAGTATTTT
5101 CTTTATGTTG AGTTAGGTTT TTCCCATCT CCGTGTAGAG GAATTTACAT
5151 ATTGTATTGG GTAAGTGTTT ACTACTTTTC CTGATTAAGG GATCTGTGCT
5201 GGGGAACAAA GCTTTTGCAG TACCTTATAT TGATGTTAAA ATTTTATTTA
5251 ACATATCCTT CAGTGAGCTC ATTTACACT GTAGCCTCTT CCTTAAATTT
5301 TGTGGTGCTC CTGTAACAGT AAGAACTAAT TCTGAAATAA AAGACATCTC
5351 CTAACAAAAA AAAAAA AAAAAA AAAAAA

```

# BLAST Results -----

Entry HS793345 from database EMBL:

human STS WI-12457.

Score = 1985, P = 1.3e-83, identities = 433/460

# Medline entries -----

97127450:

Molecular cloning of a novel 120-kDa TBP-interacting protein.

# Peptide information for frame 2 -----

ORF from 350 bp to 4039 bp; peptide length: 1230

Category: known protein  
 Classification: Nucleic acid management

```

1 MASASYHISN LLEKMTSSDK DFRFMATNDL MTELQKDSIK LDDDSERKVV
51 KMILKLEDEK NGEVQNLA VKCLGPLVSKV EYQVETIVDT LCTNMLSDEK
101 QLRDISSIGL KTVIGELPPA SSGSALAANV CKKITGRITS AIAKQEDVSV
151 QLEALDIMAD MLRSQGGLLV NFHPSILTCL LPQLTSPRLA VRKRTIIALG
201 HLVMSCGNIV FVDLIEHLLS ELSKNDMSMT TRTYIQICIA ISRQAGHRIG
251 EYLEKIIPLV VKFCNVDDDE LREYCIQAFE SFVRRCPKEV YPHVSTIINI
301 CLKYLTYPDN YNYDDEDEDE NAMDADGGDD DDQGSDEYS DDDMSWKVR
351 RAAAKCLDAV VSTRHEMLPE FYKTVSPALI SRFKEREENV KADVHFAYLS
401 LLKQTRPVQS WLCDDPDAMEQ GETPLTMLQS QVPNIVKALH KQMKESVKT
451 RQCCFNMLTE LVNVLPALT QHIPVLVPGI IFSLNDKSSS SNLKIDALSC
501 LYVILCNHSP QVFPHPVQAL VPPVAVCGD PFYKITSEAL LVTQQLVKVI
551 RPLDQSSFD ATPYKDLFT CTIKRLKAAD IDQEVKERAI SCMQIICNL
601 GDNLGSDLPN TLQIFLERLK NEITRLTTVK ALTLIAGSPL KIDLRPVLGE
651 GVPIIASFLR KNQRAKLGK LSALDILIKN YSDSLTAAMI DAVLDELPLP
701 ISESDMHVSQ MAISFLTTLA KVPSSLSKI SGSILNELIG LVRSPLLQGG
751 ALSAMLDFFQ ALVVTGTNNL GYMDLLRMLT GPVYSQSTAL THKQSYYSIA
801 KCVAALTRAC PKEGPAVVGQ FIQDVKNRSR TDSIRLLALL SLGEVGHHD
851 LSGQLELKS V ILEAFSSPSE EVKSAASYAL GSISVGNLPE YLPFVLQEI
901 SOKPRQYLL HSLKEIISA SVVGLKPYVE NIWALLKHC ECAEEGTRNV
951 VAECGLKTL IDPETLLPRL KGYLISGSSY ARSSVVTAVK FTISDHPQPI
1001 DPLLKNCIGD FLKTLDPDL NVRRVALVTF NSAAHNKPSL IRDLDTVLP
1051 HLYNETKVRK ELIREVEMGP FKHTVDDGLD IRKAAFECEY TLDSCLDRL
1101 DIFEFLNHVE DGLKDHYDIK MLTFLMLVRL STLCPASVLQ RLDRLVEPLR
1151 ATCTTKVKAN SVKQFEKQD ELKRSAMRAV AALLTIPEAE KSPLMSEFQS
1201 QISSNPELAA IFESIQKDS STNLESMDTS

```

## BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphtes3\_22g2, frame 2

TREMBL:AB020636\_1 gene: "KIAA0829"; product: "KIAA0829 protein"; Homo sapiens mRNA for KIAA0829 protein, partial cds., N = 1, Score = 5986, P = 0

TREMBL:RND6711\_1 gene: "tip120"; product: "TIP120"; Rattus norvegicus mRNA for TIP120, complete cds., N = 1, Score = 6203, P = 0

>TREMBL:RND6711\_1 gene: "tip120"; product: "TIP120"; Rattus norvegicus mRNA for TIP120, complete cds.  
 Length = 1,230

## HSPs:

Score = 6203 (930.7 bits), Expect = 0.0e+00, P = 0.0e+00  
 Identities = 1227/1230 (99%), Positives = 1228/1230 (99%)

```

Query: 1 MASASYHISN LLEKMTSSDK DFRFMATNDL MTELQKDSIK LDDDSERKVV KMILKLEDEK 60
Sbjct: 1 MASASYHISN LLEKMTSSDK DFRFMATNDL MTELQKDSIK LDDDSERKVV KMILKLEDEK 60

Query: 61 NGEVQNLA VKCLGPLVSKV EYQVETIVDT LCTNMLSDEK QLRDISSIGL KTVIGELPPA 120
Sbjct: 61 NGEVQNLA VKCLGPLVSKV EYQVETIVDT LCTNMLSDEK QLRDISSIGL KTVIGELPPA 120

Query: 121 SSGSALAANV CKKITGRITSAIAKQEDVSVQLEALDIMAD MLRSQGGLLVNFHPSILTCL 180
Sbjct: 121 SSGSALAANV CKKITGRITSAIAKQEDVSVQLEALDIMAD MLRSQGGLLVNFHPSILTCL 180

Query: 181 LPQLTSPRLA VRKRTIIALGHLVMSCGNIVFVDLIEHLLSELSKNDMSMT TRTYIQICIA 240
Sbjct: 181 LPQLTSPRLA VRKRTIIALGHLVMSCGNIVFVDLIEHLLSELSKNDMSMT TRTYIQICIA 240

Query: 241 ISRQAGHRIG EYLEKIIPLVVKFCNVDDDELREYCIQAFESFVRRCPKEV YPHVSTIINI 300
Sbjct: 241 ISRQAGHRIG EYLEKIIPLVVKFCNVDDDELREYCIQAFESFVRRCPKEV YPHVSTIINI 300

Query: 301 CLKYLTYPDN YNYDDEDEDE NAMDADGGDDDDQGSDEYS DDDMSWKVRRAAAKCLDAV 360
Sbjct: 301 CLKYLTYPDN YNYDDEDEDE NAMDADGGDDDDQGSDEYS DDDMSWKVRRAAAKCLDAV 360

Query: 361 VSTRHEMLPE FYKTVSPALISRFKEREENVKADVHFAYLSLLKQTRPVQSWL CDDPDAMEQ 420
Sbjct: 361 VSTRHEMLPE FYKTVSPALISRFKEREENVKADVHFAYLSLLKQTRPVQSWL CDDPDAMEQ 420

```

Query: 421 GETPLTMLQSQVNPINIVKALHKQMKESVKTRQCCFNMLTELNVNLPALTQHIPVLPVPGI 480  
 GETPLTMLQSQVNPINIVKALHKQMKESVKTRQCCFNMLTELNVNLPALTQHIPVLPVPGI  
 Sbjct: 421 GETPLTMLQSQVNPINIVKALHKQMKESVKTRQCCFNMLTELNVNLPALTQHIPVLPVPGI 480

Query: 481 IFSLNDKSSSSNLKIDALSCLYVILCNHSPQVFHPPHVALVPPVACVGDPPFYKITSEAL 540  
 IFSLNDKSSSSNLKIDALSCLYVILCNHSPQVFHPPHVALVPPVACVGDPPFYKITSEAL  
 Sbjct: 481 IFSLNDKSSSSNLKIDALSCLYVILCNHSPQVFHPPHVALVPPVACVGDPPFYKITSEAL 540

Query: 541 LVTQQLVKVIRPLDQPSFDPATPYIKDLFTCTIKRLKAADIDQEVKERAISCMGQIICNL 600  
 LVTQQLVKVIRPLDQPSFDPATPYIKDLFTCTIKRLKAADIDQEVKERAISCMGQIICNL  
 Sbjct: 541 LVTQQLVKVIRPLDQPSFDPATPYIKDLFTCTIKRLKAADIDQEVKERAISCMGQIICNL 600

Query: 601 GDNLGSDLPNTLQIFLERLKNEITRLTTVKALTLIAGSPKIDLRPVLGEGVPILASFLR 660  
 GDNLG DL NTLQIFLERLKNEITRLTTVKALTLIAGSPKIDLRPVLGEGVPILASFLR  
 Sbjct: 601 GDNLGSDLPNTLQIFLERLKNEITRLTTVKALTLIAGSPKIDLRPVLGEGVPILASFLR 660

Query: 661 KNQRALKGLTSLDILIKNYSDSLTAAMIDAVLDELPLPISSEDMHVSQMAISFLTTLA 720  
 KNQRALKGLTSLDILIKNYSDSLTAAMIDAVLDELPLPISSEDMHVSQMAISFLTTLA  
 Sbjct: 661 KNQRALKGLTSLDILIKNYSDSLTAAMIDAVLDELPLPISSEDMHVSQMAISFLTTLA 720

Query: 721 KVPSSLSKISGSILNELIGLVRSPLQGGALSAMLDFFQALVVTGTNNLGMDLLRMLT 780  
 KVPSSLSKISGSILNELIGLVRSPLQGGALSAMLDFFQALVVTGTNNLGMDLLRMLT  
 Sbjct: 721 KVPSSLSKISGSILNELIGLVRSPLQGGALSAMLDFFQALVVTGTNNLGMDLLRMLT 780

Query: 781 GPVYSQSTALTHKQSYYSIAKCAALTRACPKEGPAVVGQFIQDVKNRSRSTDSIRLLALL 840  
 GPVYSQSTALTHKQSYYSIAKCAALTRACPKEGPAVVGQFIQDVKNRSRSTDSIRLLALL  
 Sbjct: 781 GPVYSQSTALTHKQSYYSIAKCAALTRACPKEGPAVVGQFIQDVKNRSRSTDSIRLLALL 840

Query: 841 SLGEVGHHDLSGQLELKSIVLEAFSSPSEEVKSAASYALGSISVGNLPEYLPFVLQEIT 900  
 SLGEVGHHDLSGQLELKSIVLEAFSSPSEEVKSAASYALGSISVGNLPEYLPFVLQEIT  
 Sbjct: 841 SLGEVGHHDLSGQLELKSIVLEAFSSPSEEVKSAASYALGSISVGNLPEYLPFVLQEIT 900

Query: 901 SQPKRQYLLHSLKEIISASVVGKPYVENIALLKHCECAEEGTRNVVAECLGKLT 960  
 SQPKRQYLLHSLKEIISASVVGKPYVENIALLKHCECAEEGTRNVVAECLGKLT  
 Sbjct: 901 SQPKRQYLLHSLKEIISASVVGKPYVENIALLKHCECAEEGTRNVVAECLGKLT 960

Query: 961 IDPETLLPRLKGYLISGSSYARSSVVTAVKFTISDHPQPIDPLKNCIGDFLKTLEDPLD 1020  
 IDPETLLPRLKGYLISGSSYARSSVVTAVKFTISDHPQPIDPLKNCIGDFLKTLEDPLD  
 Sbjct: 961 IDPETLLPRLKGYLISGSSYARSSVVTAVKFTISDHPQPIDPLKNCIGDFLKTLEDPLD 1020

Query: 1021 NVRRVALVTFNSAAHNKPSLIRDLDTVLPHLYNETKVRKELIREVEMGPFKHTVDDGLD 1080  
 NVRRVALVTFNSAAHNKPSLIRDLDTVLPHLYNETKVRKELIREVEMGPFKHTVDDGLD  
 Sbjct: 1021 NVRRVALVTFNSAAHNKPSLIRDLDSVLPHLYNETKVRKELIREVEMGPFKHTVDDGLD 1080

Query: 1081 IRKAAFECCMYTLLDSCDLRLDIFEFLNHVEDGLKDHYDIKMLTFLMLVRLSTLCPSAVLQ 1140  
 IRKAAFECCMYTLLDSCDLRLDIFEFLNHVEDGLKDHYDIKMLTFLMLVRLSTLCPSAVLQ  
 Sbjct: 1081 IRKAAFECCMYTLLDSCDLRLDIFEFLNHVEDGLKDHYDIKMLTFLMLVRLSTLCPSAVLQ 1140

Query: 1141 RLDRLVEPLRATCTTKVKANSVKQEFQKDELKRSAMRAVAALLTIPEAEKSPLMSEFQS 1200  
 RLDRLVEPLRATCTTKVKANSVKQEFQKDELKRSAMRAVAALLTIPEAEKSPLMSEFQS  
 Sbjct: 1141 RLDRLVEPLRATCTTKVKANSVKQEFQKDELKRSAMRAVAALLTIPEAEKSPLMSEFQS 1200

Query: 1201 QISSNPELAAIFESIQRDSSSTNLESMDTS 1230  
 QISSNPELAAIFESIQRDSSSTNLESMDTS  
 Sbjct: 1201 QISSNPELAAIFESIQRDSSSTNLESMDTS 1230

Pedant information for DKF2phtes3\_22g2, frame 2

# Report for DKF2phtes3\_22g2.2

[LENGTH] 1230  
 [MW] 136376.58  
 [pI] 5.52  
 [HOMOL] TREMBL:RND6711\_1 gene: "tip120"; product: "TIP120"; Rattus norvegicus mRNA for  
 TIP120, complete cds. 0.0  
 [KW] TRANSMEMBRANE 1  
 [KW] LOW\_COMPLEXITY 5.28 %

SEQ MASASYHISNLEKMTSSDKDFRFRMATNDLMTLQKDSIKLDDDSERKVVKMILKLEDEK  
 SEG .....  
 PRD cccccchhhhhhhhhccccceeeehhhhhhhhhccccccccchhhhhhhhhhhhhcc  
 MEM .....  
 SEQ NGEVQNLAVKCLGPLVSKVKEYQVETIVDTLCTNMLSDKEQLRDISSIGLKTIVIGELPPA  
 SEG .....xxxx  
 PRD cccccceeeeeeceeeehhhhhhhhhhhccccchhhhhccccccccchhhhhhhhhcccc

MEM .....  
SEQ SSGSALAANVCKKITGRLTSAIAKQEDVSVQLEALDIMADMLSRQGGLLVNFHPSILTCL  
SEG xxxxxxxx .....  
PRD cccccchhhhhhhccchhhhhhhccccchhhhhhhhhhhhhhhccceeeccchhhhh  
MEM .....  
SEQ LPQLTSPRLAVKRRTIIALGHLVMSGNI FVDLIEHLLSELSKNDSMSTTRTYIQCIAA  
SEG .....  
PRD hccccchhhhhhhhhhhheeeccccceehhhhhhhhhhhcccccchhhhhhhhhhh  
MEM MMMMMMMMMMMMMMM .....  
SEQ ISRQAGHRIGEYLEKIIPLVVKFCNVDDDELREYCIQAFESFVRRCPEVYPHVSTIINI  
SEG .....  
PRD hhhccccccccchhhhhhhheeeccchhhhhhhhhhhhhhhccccceecchhhhh  
MEM .....  
SEQ CLKYLTYDPNINYDDEDEDENAMDADGGDDDDQGSDDDEYSDDDMSWKVRAAAKCLDAV  
SEG ..... xxx .....  
PRD hhhhhccccccccccccccccccccccccccccccccccccchhhhhhhhhhhhhhh  
MEM .....  
SEQ VSTRHEMLPEFYKTVSPALISRFKEREENVKADVHFAYLSLLKQTRPVQSWLCDPDAMEQ  
SEG .....  
PRD hhhhhhhhhhhhhhhhhchhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhccceeeccccccc  
MEM .....  
SEQ GETPLTMLQSQVPNI V KALHKQMEKSVKTRQCCFNMLTELNVNVLPGALTQHIPVLVPGI  
SEG .....  
PRD cccccchhhccccccccceeeccce  
MEM .....  
SEQ IFSLNDKSSSNL KIDALSCLYVILCNHSPQVFHPHVQALVPPVAVCGDPFYKITSEAL  
SEG xxxxxxxxxxxxxxxx .....  
PRD eeeccccccccchhhhhhhheeeccccccccccccceeeccccceccccchhhhhhh  
MEM .....  
SEQ LVTQQLVKVIRPLDQPSFDPATPYIKDLFTCTIKRLKAADIDQEVKERAISCMGQIICNL  
SEG .....  
PRD hhhhhhhhhhhccccccccccccccccchhhhhhhhhhhhhhhhhhhhhhhhhhhheeecc  
MEM .....  
SEQ GDNLGSDLPNTLQIFLERLKNEITRLTTVKALTLIAGSPLKIDLRPVLGEGVPI LASFLR  
SEG .....  
PRD cccccccccchhhhhhhhhchhhhhhhhhhhhhheeeccccccccceehhhhhhhhhhh  
MEM .....  
SEQ KNQRALKGLTSLADILIKNYSDSLTAAMIDAVLDELPLISESDMHVSQMAISFLTTLA  
SEG .....  
PRD hhhhhhhhhhhhhhhhhhhhhccccchhhhhhhhhhhccccccccchhhhhhhhhhhhhhh  
MEM .....  
SEQ KVYPSSLSKISG SILNELIGLVRSPLLQGGALSAMLDFQALVVTGTNNLG YMDLLRMLT  
SEG .....  
PRD cccccceecchhhhhhhhhhhccccccccchhhhhhhhhhhheeeccccchhhhhhhhhc  
MEM .....  
SEQ GPVYSQSTALTHKQSYYSIAKCVAAALTRACPKEGPAVVGGQFIQDVKNRSTDSIRLLALL  
SEG .....  
PRD cccccccccchhhhhhhhhhhhhhhhhhhhhhhccccchhhhhhhhhhhccccchhhhhhh  
MEM .....  
SEQ SLGEVGHHDLSGQLELKS VILEAFSSPSEEVKSAASYALGSISVGNLPEYLPFVLQEIT  
SEG .....  
PRD hccccccccccccccccceeeccccchhhhhhhhhhhccccccccccccchhhhhhhhh  
MEM .....  
SEQ SQPKRQYLLLSLKETISSASVVG LKPYVENI W ALLKHCECAEEGTRNVVAECLGKLT  
SEG .....  
PRD cccchhhhhhhhhhhhhccccceehhhhhhhhhhhhhhhhhccccceeecccccccc  
MEM .....  
SEQ IDPETLLPRLKGYLISGSSYARSSVVTA VKFTISDHPQPIDPLLKNCIGDFLKTLED PDL  
SEG .....  
PRD cccccccccccccccccccccchhhhhhhhhhhhhccccccccccccchhhhhhhhhhhccccc  
MEM .....  
SEQ NVRVALVTFN SAAHNKPSLIRDLLDTVLPHLYNETKVRKELIREVEMGPFKHTVDDGLD  
SEG .....  
PRD cceeeeeeccccccccchhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhccccccccccccch  
MEM .....

(No Pfam data available for DKFZphtes3\_22g2.2)



DKFZphtes3\_22n13

group: testes derived

DKFZphtes3\_22n13 encodes a novel 677 amino acid protein without similarity to known proteins.

No informative BLAST results; No predictive prosite, pfam or SCOP motif.

The new protein can find application in studying the expression profile of testis-specific genes.

dJ1042K10.3, complete

Sequenced by LMU

Locus: /map="22q13.1-13.2"

Insert length: 3353 bp

Poly A stretch at pos. 3315, polyadenylation signal at pos. 3298

```
1 ATGGAACCCAC TATCCCCACT GCCAAGTCCA CCCCACACT CATTAGCAA
51 AGCCAAACCCA AGTCTGCCAG TGAGAAGTCA CAGCGCAGCA AGAAGGCCAA
101 GGAGCTGAAG CCAAAGGTGA AGAAGCTCAA GTACCACCAG TACATCCCCC
151 CGGACCAGAA GCAGGACAGG GGGGCACCCC CCATGGACTC ATCCTACGCC
201 AAGATCCTGC AGCAGCAGCA GCTCTTCTC CAGCTGCAGA TCCTCAACCA
251 GCAGCAGCAG CAGCACCACA ACTACCAGGC CATCCTGCCT GCCCGCCAA
301 AGTCAGCAGG CGAGGCCCTG GGAAAGCAGC GGACCCCCC AGTACGCAGC
351 CTCTCCACTA CCAATAGCAG CTCCAGCTCG GGCGCCCTG GGCCCTGTGG
401 GCTGGCAGCT CAGAACAGCA CCTCACTGAC TGGCAAGCCG GGAGCCCTGC
451 CGGCCAACCT GGACGACATG AAGGTGGCAG AGCTGAAGCA GGAGCTGAAG
501 TTGCGATCAC TGCTGTCTC GGGCACCAA ACTGAGCTGA TTGAGCGCCT
551 TCGAGCCTAT CAAGACCAA TCAGCCCTGT GCCAGGAGCC CCAAGGCCC
601 CTGCGCCGAC CTCTATCCTG CACAAGGCTG GCGAGGTGGT GGTAGCCTTC
651 CCAGCGGCCG GGCTGAGCAC GGGGCCAGCC CTGGTGGCAG CAGGCCCTGGC
701 TCCAGCTGAG GTGGTGGTGG CCACGGTGGC CAGCAGTGGG GTGGTGAAGT
751 TTGGCAGCAC GGGCTCCAGC CCCCCCGTGT CTCCACCCCC CTCGGAGCGC
801 TCAGTGTCTA GCACGGGCGA TGAAAACTCC ACCCCCGGGG ACACCTTTGG
851 TGAGATGGTG ACATCACCTC TGACGCAGCT GACCCTGCAG GCCTCGCCAC
901 TCGAGATCCT CGTGAAGGAG GAGGGCCCCC GGGCCGGGTC CTGTTGCCCTG
951 AGCCCTGGGG GGGCGGCGGA GCTAGAGGGG CGGACAAGG ACCAGATGCT
1001 GCAGGAGAAA GACAAGCAGA TCGAGGCGCT GACGCGCATG CTCGGGCGA
1051 AGCAGCAGCT GGTGGAGCGG CTCAGCTGTC AGCTGGAGCA GGAGAAGCGA
1101 GCCCAGCAGC CCGCCCCCGC CCCCCCCCCC CTCGGCACCC CCGTGAAGCA
1151 GGAGAACAGC TTCTCCAGCT GCCAGCTGAG CCAGCAGCCC CTGGGCCCGC
1201 CTCACCCATT CAACCCAGC CTGGCGGCC CAGCCACCAA CCACATAGAC
1251 CTTTGTGCTG TGGCCCCAGG GCCCCCGTCC GTGGTGGTGA AGCAGGAAGC
1301 CTTGCAGCCT GAGCCCCGAG CGGTCCCCGC CCCCCAGTTG CTTCTGGGCG
1351 CTCAGGGCCC CGGCCTCATC AAGGGGGTTG CACCTCCAC CCTCATCACC
1401 GACTCCACAG GGACCCACCT TGTCTCACG GTGACCAATA AGAATGCAGA
1451 CAGCCCTGGC CTGTCCAGTG GGAGCCCCCA GCAGCCCTCG TCCCAGCCTG
1501 GCTCTCCAGC GCCTGCCCCC TCTGCCAGA TGGACCTGGA GCACCCACTG
1551 CAGCCCTCTT TTGGGACCCC CACTTCTCTG CTGAAGAAGG AACCACCTGG
1601 CTATGAGGAA GCCATGAGCC AGCAGCCCAA ACAGCAGGAA AATGGTTCTT
1651 CAAGCCAGCA GATGGACGAC CTGTTTGACA TTCTCATTCA GAGCGGAGAA
1701 ATTTACAGCAG ATTTCAAGGA GCCGCCATCC CTGCCAGGGA AGGAGAAGCC
1751 ATCCCCGAAG ACAGTCTGTG GGTCCCCCCT GGCAGCACAG CCATCACCTT
1801 CTGCTGAGCT CCCCCAGGCT GCCCCACCTC CTCCAGGCTC ACCCTCCCTC
1851 CCTGGAGCCC TGGAGGACTT CCTGGAGAGC AGCACGGGGC TGCCCTTGCT
1901 GACCAGTGGG CATGACGGGC CAGAGCCCTT TTCCCTCATT GACGACCTCC
1951 ATAGCCAGAT GCTGAGCAGC ACTGCCATCC TGGACCACCC CCGTCACCCC
2001 ATGGACACCT CGGAATTGCA CTTTGTTCCT GAGCCCAGCA GCACCATGGG
2051 CCTGGACCTG GCTGATGGCC ACCTGGACAG CATGGACTGG CTGGAGCTGT
2101 CGTCAGGTGG TCCCGTGTCT AGCCTAGCCC CCCTCAGCAC CACAGCCCCC
2151 AGCCTCTTCT CCACAGACTT CCTCGATGGC CATGATTTCG AGCTGCACTG
2201 GGATTCTCTG TTGTAGCTCT CTGGCTCAAG ACGGGGTGGG GAAGGGGCTG
2251 GGAGCCAGGG TACTCCAATG CGTGGCTCTC CTGCGTGATT CGGCCCTCTC
2301 ACATGGTTGT GAGTCTTGAC AATCACAGCC CTGCTTTT CTCTCCCTG
2351 GGAGGCTAGA ACAGAGAAGC CTTTACTCCT GGTTCAGTGC CACGCAGGGC
2401 AGAGGAGAGC AGCTGTCAAG AAGCAGCCCT GGCTCTCACG CTGGGGTTTT
2451 GGAGACACAG TCAGGGTCAG GGCCATTTCA GCTTGACCTC CTTTTTTGAG
2501 GTCAGGGGGC ACTGCTGTCT TGGCTACAAT TTGGCTAAGG TAGGTGAAGC
2551 CTGGCCAGGC GGGAGGCTTC TCTTCTGACC CAGGGCTGAG ACAGGTTAAG
2601 GGGTGAATCT CCTTCCTTTC TCTCCCTGCT TTGCTGTGAA GGGAGAAATT
2651 AGCCTGGGCC TCTACCCCTT ATTCCCTGTG TCTGCCAACC CCAGGATCCC
2701 AGGGCTCCCT GCCATTTTAG TGTCTTGGTG TAGTGTAACC ATTTAGTGGT
2751 TGGTGGCAAC AATTTTATGT ACAGGTGTAT ATACCTCTAT ATTATATATC
2801 GACATACATA TATATTTTGG GGGGGGGGCG GACAGGAGAT GGGTGCAACT
```

```

2851 CCCTCCCATC CTA CTCTCAC AGAAGGGCCT GGATGCAAGG TTACCCCTGA
2901 GCTGTGTGCC ACAGTCTGGT GCCCAGTCTG GCATGCAGCT ACCCAGGCCC
2951 ACCCATCACG TGTGATTGAC ATGTAGGTAC CCTGCCACGG CCTATGCCCC
3001 ACCTGCCCTG CTTCTGGCT CTTATCAGT GCCATGAGGG CAGAGGTGCT
3051 ACCTGGCCTT CCTGCCAGGA GCTCTCCACC CACTCACATT CCGTCCCCGC
3101 CGCCTCACTG CAGCCAGCGT GGCCTTAGGA CAGGAGGAGC TTCGGGCCCCA
3151 GCTTCACCTT GCGGTGGGGC TGAGGGGTGG CCATCTCCTG CCTGGGGCC
3201 ACTGGCTTCA CATTCTGGGC TGA CTATAG GGGAGTAGGG GTGGAGTCAC
3251 CAAAACCACT GCTGGGACAA AGATGGGGAA GGTGTGTGAA CTTTTTAAAA
3301 TAAACACAAA AACACAGGAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA
3351 AAG

```

## BLAST Results

Entry HS1042K10 from database EMBL:  
 Human DNA sequence from clone 1042K10 on chromosome 22q13.1-13.2.  
 Contains the ADSL gene for Adenylosuccinate lyase (EC 4.3.2.2,  
 Adenylosuccinase, ASL) and 4 novel genes (one with probable rabGAP  
 domains and Src homology domain 3). Contains ESTs, STSs, GSSs and a  
 putative CpG island.  
 Score = 7997, P = 0.0e+00, identities = 1617/1645  
 7 exons

## Medline entries

No Medline entry

## Peptide information for frame 3

ORF from 183 bp to 2213 bp; peptide length: 677  
 Category: similarity to unknown protein  
 Classification: unclassified

```

1 MDSSYAKILQ QQQLFLQLQI LNQQQQQHNN YOAILPAPPK SAGEALGSSG
51 TPPVRSLSST NSSSSSGAPG PCGLARQNST SLTGKPGALP ANLDDMKVAE
101 LKQELKLRSL PVSGTKTELI ERLRAYQDQI SPVPGAPKAP AATSILHKAG
151 EVVVAFFPAAR LSTGPALVAA GLAPAEVVVA TVASSGVVVF GSTGSTPPVS
201 PTPSERSLLS TGDENSTPGD TFGEMVTSPL TQLTLOASPL QILVKEEGPR
251 AGSCCLSPGG RAELEGRDKD QMLQEKDKQI EALTRMLRQK QQLVERLKLQ
301 LEQEKRAQPP APAPAPLGTP VKQENSFSSC QLSQQPLGPA HPFNPSLAAP
351 ATNHIDPCAV APGPPSVVVK QEALQPEPEP VPAPQLLLGP QGPGLIKGVA
401 PPTLITDSTG THLVLTVINK NADSPGLSSG SPQQPSSQPG SPAPAPSAQM
451 DLEHPLQLPF GTPTSLLKKE PPGYEEAMSQ QPKQENGSS SQQMDLFDI
501 LIQSGEISAD FKEPPSLPGK EKPSPKTVCG SPLAAQSPS AELPQAAPP
551 PGSPSLPGRL EDFLESSTGL PLLTSGHDGP EPLSLIDDLH SQMLSSTAIL
601 DHPPSPMDTS ELHFVPEPSS TMGLDLADGH LDSMDWLELS SGPVLSLAP
651 LSTTAPSLFS TDFLDGHDQ LHWDSCL

```

## BLASTP hits

No BLASTP hits available

## Alert BLASTP hits for DKFZphtes3\_22n13, frame 3

TREMBL:HS1042K10\_6 gene: "dJ1042K10.3"; product: "dJ1042K10.3 (novel protein)"; Human DNA sequence from clone 1042K10 on chromosome 22q13.1-13.2. Contains the ADSL gene for Adenylosuccinate lyase (EC 4.3.2.2, Adenylosuccinase, ASL) and 4 novel genes (one with probable rabGAP domains and Src homology domain 3). Contains ESTs, STSs, GSSs and a putative CpG island., N = 1, Score = 1285, P = 4.9e-131

TREMBL:CEUK06A9\_3 gene: "K06A9.1a"; Caenorhabditis elegans cosmid K06A9., N = 2, Score = 149, P = 1.3e-09

TREMBLNEW:SSI132828\_1 product: "p210 protein"; Spermatozopsis similis mRNA for p210 protein, partial, N = 1, Score = 171, P = 2.8e-09

>TREMBL:HS1042K10\_6 gene: "dJ1042K10.3"; product: "dJ1042K10.3 (novel protein)"; Human DNA sequence from clone 1042K10 on chromosome 22q13.1-13.2. Contains the ADSL gene for Adenylosuccinate lyase (EC

**HSPs :**

Query:	435	PSQQPGSPAPAPSAQMDLEHPLQLPLFGTPTSLLLKKEPPGYEAMSSQQPKQENGSSSSQOM	494
Sbjct:	1	PSQQPGSPAPAPSAQMDLEHPLQLPLFGTPTSLLLKKEPPGYEAMSSQQPKQENGSSSSQOM	60
Query:	495	DDLFDILIQSGEISADFKEPPSLPGKEKPSPKTVCGSPLAAQSPSAELPQAAPPPPGSP	554
Sbjct:	61	DDLFDILIQSGEISADFKEPPSLPGKEKPSPKTVCGSPLAAQSPSAELPQAAPPPPGSP	120
Query:	555	SLPGRLEDFLESSTGLPLLTSGHDGPEPLSLIDDLHSQMLSSTAILDHPPSPMDTSELHF	614
Sbjct:	121	SLPGRLEDFLESSTGLPLLTSGHDGPEPLSLIDDLHSQMLSSTAILDHPPSPMDTSELHF	180
Query:	615	VPEPSSMTGLDLADGHLDSMDWLELSSGGPVLSLAPLSTTAPSLFSTDFLDGHDQLQLHWD	674
Sbjct:	181	VPEPSSMTGLDLADGHLDSMDWLELSSGGPVLSLAPLSTTAPSLFSTDFLDGHDQLQLHWD	240
Query:	675	SCL 677	
Sbjct:	241	SCL 243	

Pedant information for DKFZphtes3\_22n13, frame 3

Report for DKFZphtes3\_22n13.3

```

[LENGTH]      677
[MW]           70743.01
[pI]           4.93
[HOMOL]        TREMBL:HS1042K10_6 gene: "dJ1042K10.3"; product: "dJ1042K10.3 (novel protein)";
Human DNA sequence from clone 1042K10 on chromosome 22q13.1-13.2. Contains the ADSL gene for
Adenylosuccinate lyase (EC 4.3.2.2, Adenylosuccinase, ASL) and 4 novel genes (one with
probable rabGAP domains and Src homology domain 3). Contains ESTs, STSs, GSSs and a putative
CpG island. 1e-111
[KW]           TRANSMEMBRANE 1
[KW]           LOW_COMPLEXITY 21.57 %
[KW]           COILED_COIL 4.58 %

```

731

```
SEG      .....xxxxxxxxx.....
PRD      hhhhhhhhcccccccccccccccccccccccccccccccccccccccccccccccccc
COILS    CCCCCC.....
MEM      .....

SEQ      APGPPSVVVKQEQALQPEPEVPVAPQQLLGPGGGLIKGVAPPTLITDSTGTHLVLTVTK
SEG      .....xxxxxxxxx.....
PRD      cccccceeecccccccccccccccccccccccccccccccccccccccccccccccc
COILS    .....
MEM      .....

SEQ      NADSPGLSSGSPQPSSQPGSPAPAPSAQMDLEHPLQLPLFGTPTSLLKKEPPGYEEAMSQ
SEG      .....xxxxxxxxx.....
PRD      ccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc
COILS    .....
MEM      .....

SEQ      QPKQKENGSSSQMDDLFDILIQSGEISADFKEPPSLPGKEKPSPKTVCGSPLAAQPSPS
SEG      .....xxxxxxxxx.....
PRD      ccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc
COILS    .....
MEM      .....

SEQ      AELPQAAPPPPGSPSLPGRLEDFLESSTGLPLLTSGHDGPEPLSLIDDLHSQMLSSTAIL
SEG      .....xxxxxxxxx.....
PRD      ccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc
COILS    .....
MEM      .....

SEQ      DHPPSPMDTSELHFVPEPSSTMGLDLADGHLDSMDWLELSSGGPVLSLAPLSTTAPSLFS
SEG      .....
PRD      ccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc
COILS    .....
MEM      .....

SEQ      TDFLDGHDLQLHWDSC
SEG      .....
PRD      cccccceeecccccc
COILS    .....
MEM      .....
```

(No Prosite data available for DKFZphtes3\_22n13.3)

(No Pfam data available for DKFZphtes3\_22n13.3)

DKFZphtes3\_23111

group: intracellular transport and trafficking

DKFZphtes3\_23111 encodes a novel 186 amino acid protein nearly identical to mouse ADP-ribosylation-like factor homolog 6 (Arl6).

Protein secretion through the endoplasmic reticulum and the Golgi vesicular trafficking system is initiated by the binding of ADP-ribosylation factors (ARFs) to donor membranes, leading to recruitment of cocatomer, bud formation, and eventual vesicle release. ARFs are approximately 20-kDa GTPases that are active with bound GTP and inactive with GDP bound. The novel protein contains an ATP/GTP-binding site motif A (P-loop) and seems to be a novel ARF. It seems to have an important role in vesicular transport and vesicular trafficking.

The new protein can find application in modulating vesicle transport and trafficking in cells.

nearly identical to mouse Arl6, ADP-ribosylation-like factor homolog

start at Bp 15 matches kozak consensus ANNAtgG

Sequenced by LMU

Locus: unknown

Insert length: 717 bp

Poly A stretch at pos. 689, no polyadenylation signal found

```
1 ATTTGAATCA CATTATGGGA TTGCTAGACA GACTTTCAGT CTTGCTTGGC
51 CTGAAGAAGA AGGAGGTTC TGTTTTGTGC CTTGGGCTAG ATAATAGTGG
101 CAAAACGACG ATCATTAAAC AACTTAAACC TTCAAATGCT CAATCTCAAA
151 ATATCCTTCC AACAAATAGGA TTCAGCATAG AGAAATTCAA ATCATCCAGT
201 TTGTCAATTA CAGTGTTTGA CATGTCAGGT CAAGGAAGAT ACAGAAATCT
251 CTGGGAACAC TATTATAAAG AAGGCCAAGC TATTATTTT GTCATTTGATA
301 GTAGTGATAG ATTAAGAATG GTTGTGGCCA AAGAAGAAGT CGATACTCTT
351 CTGAATCATC CAGATATTAA ACACCGTCGA ATTCCAATCT TATTCTTTGC
401 AAATAAAATG GATCTTAGAG ATGCAGTGAC ATCTGTAAAA GTGTCTCAGT
451 TGCTGTGTTT AGAGAATATC AAAGATAAAC CCTGGCATAT TTGTGCTAGT
501 GATGCCATAA AAGGAGAAGG CTTGCAAGAA GGTGTAGACT GGCTTCAAGA
551 TCAGATCCAG ACTGTGAAGA CATGAAAAGA TAATAGTTGG AAACCTCAGC
601 AATTTTCAAT TCAAGGAATC TATCTAAGAC AAATAGAATA CATTTTGTAA
651 AAGATGTTTA TGCATCAAAA AATATAATT TCTGCTTGCA AAAAAAAAAA
701 AAAAAAAAAA AAAAAAG
```

## BLAST Results

No BLAST result

## Medline entries

No Medline entry

## Peptide information for frame 3

ORF from 15 bp to 572 bp; peptide length: 186  
Category: strong similarity to known protein  
Classification: Intracellular transport and traffic  
Prosite motifs: ATP\_GTP\_A (24-32)

```
1 MGLLDRLSVL LGLKKKEVHV LCLGLDMSGK TTIINKLKPS NAQSQNILPT
51 IGFSIEKFKS SSLSFTVFDM SQQGRYRNW EHYKQEQAI IFVIDSSDRL
101 RMVVAKEELD TLLNHPDIKH RRIPILFFAN KMDLRDAVTS VKVSQLLCLE
151 NIKDKPWHIC ASDAIKGEGL QEGVDWLQDQ IQTVKT
```

## BLASTP hits

No BLASTP hits available

## Alert BLASTP hits for DKFZphtes3\_23111, frame 3

TREMBL:AF031903.1 gene: "Arl6"; product: "ADP-ribosylation-like factor homolog ARL6"; Mus musculus ADP-ribosylation-like factor homolog ARL6 (Arl6) mRNA, complete cds., N = 1, Score = 923, P = 1.1e-92

TREMBL:CEC38D4.5 gene: "C38D4.8"; Caenorhabditis elegans cosmid C38D4, N = 1, Score = 418, P = 3.6e-39

PIR:S66337 ADP-ribosylation factor 1 - Chlamydomonas reinhardtii, N = 1, Score = 373, P = 2.1e-34

SWISSPROT:ARF1\_CHLRE ADP-RIBOSYLATION FACTOR 1., N = 1, Score = 372, P = 2.7e-34

>TREMBL:AF031903.1 gene: "Arl6"; product: "ADP-ribosylation-like factor homolog ARL6"; Mus musculus ADP-ribosylation-like factor homolog ARL6 (Arl6) mRNA, complete cds.  
Length = 186

## HSPs:

Score = 923 (138.5 bits), Expect = 1.1e-92, P = 1.1e-92  
Identities = 178/186 (95%), Positives = 184/186 (98%)

Query: 1 MGLLDRLSVLLGLKKKEVHVLCGLDMSGKTTIINKLKPSNAQSQNILPTIGFSIEKFKS 60  
MGLLDRLS LLGLKKKEVHVLCGLDMSGKTTIINKLKPSNAQSQ+I+PTIGFSIEKFKS  
Sbjct: 1 MGLLDRLSGLLGLKKKEVHVLCGLDMSGKTTIINKLKPSNAQSQDIVPTIGFSIEKFKS 60

Query: 61 SSLSFTVFDMSGQGRYRNLEWHYYKQGAIIFVIDSSDLRMVVAKEELDTLLNHPDIKH 120  
SSLSFTVFDMSGQGRYRNLEWHYYK+GQAIIFVIDSSD+LRMVVAKEELDTLLNHPDIKH  
Sbjct: 61 SSLSFTVFDMSGQGRYRNLEWHYYKDGQAIIFVIDSSDKLRMVVAKEELDTLLNHPDIKH 120

Query: 121 RRIPILFFANKMDLRDAVTSVKVSQLLCLENIKDKPWHICASDAIKGEGLQEGVDWLQDQ 180  
RRIPILFFANKMDLRD+VTSVKVSQLLCLE+IKDKPWHICASDAIKGEGLQEGVDWLQDQ  
Sbjct: 121 RRIPILFFANKMDLRDSVTSVKVSQLLCLESIKDKPWHICASDAIKGEGLQEGVDWLQDQ 180

Query: 181 IQTVKT 186  
IQ VKT  
Sbjct: 181 IQAVKT 186

## Pedant information for DKFZphtes3\_23111, frame 3

## Report for DKFZphtes3\_23111.3

[LENGTH] 186  
[MW] 21097.69  
[PI] 8.72  
[HOMOL] TREMBL:AF031903.1 gene: "Arl6"; product: "ADP-ribosylation-like factor homolog ARL6"; Mus musculus ADP-ribosylation-like factor homolog ARL6 (Arl6) mRNA, complete cds. 4e-94

[FUNCAT] 30.08 organization of golgi [S. cerevisiae, YDL192w] 1e-36  
[FUNCAT] 06.10 assembly of protein complexes [S. cerevisiae, YDL192w] 1e-36  
[FUNCAT] 08.07 vesicular transport (golgi network, etc.) [S. cerevisiae, YDL192w] 1e-36

[FUNCAT] 30.09 organization of intracellular transport vesicles [S. cerevisiae, YDL137w] 2e-36

[FUNCAT] 06.07 protein modification (glycosylation, acylation, myristylation, palmitoylation, farnesylation and processing) [S. cerevisiae, YBR164c] 2e-32  
[FUNCAT] 30.03 organization of cytoplasm [S. cerevisiae, YBR164c] 2e-32  
[FUNCAT] 03.22 cell cycle control and mitosis [S. cerevisiae, YMR138w] 4e-19  
[FUNCAT] 30.04 organization of cytoskeleton [S. cerevisiae, YMR138w] 4e-19  
[FUNCAT] x general function prediction [M. jannaschii, MJ1339] 2e-05  
[FUNCAT] 30.02 organization of plasma membrane [S. cerevisiae, YHR005c] 4e-05  
[FUNCAT] 03.07 pheromone response, mating-type determination, sex-specific proteins [S. cerevisiae, YHR005c] 4e-05

[FUNCAT] 10.05.07 g-proteins [S. cerevisiae, YHR005c] 4e-05  
[FUNCAT] 08.13 vacuolar transport [S. cerevisiae, YKR014c] 2e-04  
[FUNCAT] 08.19 cellular import [S. cerevisiae, YKR014c] 2e-04  
[FUNCAT] 06.04 protein targeting, sorting and translocation [S. cerevisiae, YKR014c] 2e-04

[FUNCAT] 03.04 budding, cell polarity and filament formation [S. cerevisiae, YFL005w] 4e-04

[BLOCKS] BL01288C  
[BLOCKS] BL01020C SAR1 family proteins  
[BLOCKS] BL01019C ADP-ribosylation factors family proteins

```

[BLOCKS]      BL01019B ADP-ribosylation factors family proteins
[BLOCKS]      BL01019A ADP-ribosylation factors family proteins
[SCOP]        dias3_2 3.29.1.4.12 Transducin (alpha subunit), insertion domain 2e-45
[SCOP]        d1mhl_ 3.29.1.4.2 Rac1 [Human (Homo sapiens) 2e-46
[SCOP]        d5p2l_ 3.29.1.4.1 cH-p21 Ras protein [human (Homo sapiens) 5e-37
[SCOP]        dihora_ 3.29.1.4.8 ADP-ribosylation factor 1 (ARF1) [human (Homo sapiens) 4e-61
[SCOP]        dla2kc_ 3.29.1.4.5 Ran Nuclear transport factor-2 (NTF2) [Do 4e-33
[PIRKW]       glycoprotein 2e-33
[PIRKW]       monomer 3e-31
[PIRKW]       P-loop 2e-35
[PIRKW]       lipoprotein 2e-33
[PIRKW]       GTP binding 2e-35
[SUPFAM]      ADP-ribosylation factor 2e-35
[PROSITE]     ATP_GTP_A 1
[PFAM]        ADP-ribosylation factors (Arf family) (contains ATP/GTP binding P-loop)
[KW]          Alpha_Beta
[KW]          3D
[KW]          LOW_COMPLEXITY 5.91 %

```

```

SEQ  MGLLDRLSVLLGLKKKEVHVLCGLDLSGKTTIINKLKPSNAQSQNILPTIGFSIEKFKS
SEG  ..xxxxxxxxxxxxx
lhurA .....CCCCEEEEETTTTCHHHHHHHHCCCCEEEE--EEETEEEEEEEE

SEQ  SLSFTVFDMSGQGRYRNLEWHYYKEGQAIIFVIDSSDRLRMVVAKEELDTLLNHPDIKH
SEG  .....
lhurA TTEEEEEETTTTTTCCCHHHHHHCEEEEEETTTTHHHHHHHHHHHHHHTTTT--

SEQ  RRIPILFFANKMDLRDAVTSVKVSQLCLLENIKDKPWHICASDAIKGEGLQEGVDWLQDQ
SEG  .....
lhurA TTTEEEEEETTTTTTCCCHHHHHHHHCGGGTTTTCEEEECBTBTBHHHHHHHHHHHH

SEQ  IQTVKT
SEG  .....
lhurA HHHHC.

```

## Prosites for DKFZphtes3\_23111.3

PS00017 24->32 ATP\_GTP\_A PDOC00017

## Pfam for DKFZphtes3\_23111.3

```

HMM_NAME      ADP-ribosylation factors (Arf family) (contains ATP/GTP binding P-loop)
HMM            *GMgWfsIFrkmWGLWNKEMRILMLGLDNAGKTTIILYMLKlgE..IVTTI
               MG++ ++ ++GL +KE+++L LGLDN+GKTTI+++LK+ ++
Query          1 -MGLLDRLSVLLGLKKKEVHVLCGLDLSGKTTIINKLKPSNAQSQNIL 48

HMM            PTIGFNVETVeYKNIKFNVDVGGQdsIRPYWRHYYPNTDGLIIVVDSaD
               PTIGF +E+ + ++F+V+D GQ + R +W HYY + ++II+V+DS+D
Query          49 PTIGFSIEKFSSLSFTVFDMSGQGRYRNLEWHYYKEGQAIIFVIDSSD 98

HMM            RDRMeEaKqELHaMLNEEEL..rDAPILIFANKQDLPgAMSeSIEIREaLG
               R RM AK+EL+ +LN+ ++ R+ P+L FANK DL++A+++ +++ +L
Query          99 RLRMVVAKEELDTLLNHPDIKHRRIPILFFANKMDLRDAVTSVKVSQLLC 148

HMM            LHeIRCnRPWYIQMCCAvtGEGLYEGMDWLSNYInkrkK*
               L++I+ + PW+I +++A++GEG+EG DWL ++I+ K
Query          149 LENIK-DKPWHICASDAIKGEGLQEGVDWLQDQIQTVKT 186

```

DKFZphtes3\_23n19

group: testes derived

DKFZphtes3\_23n19 encodes a novel 387 amino acid protein with similarity to rat protein kinase C-interacting RBCC protein 1.

The novel protein contains not the RING-B box-coiled coil (RBCC) motif of RBCC protein 1, and thus is not a member of this subgroup of RING finger proteins.  
No informative BLAST results; No predictive prosite, pfam or SCOP motifs.

The new protein can find application in studying the expression profile of testis-specific genes.

similarity to rat protein kinase C-interacting RBCC protein 1

start at Bp 209 matches kozak consensus PyNNatgG  
similarity to of C-terminal part to N-terminus of RBCK1

Sequenced by LMU

Locus: unknown

Insert length: 1579 bp

Poly A stretch at pos. 1535, polyadenylation signal at pos. 1515

```
1 CGGAGACCCT CGGGCCGTGT CCATTGTGG GCAAAGCCAG CGGGGCAGGC
51 TTGGCCAGAG TGCACCACTC GCGCCCGTCC CAGGCCCGAC GCTCTGGGCG
101 CGCCCCGAAC CCCAGGTTTCG CGGCCCGTGT TTCCGACCGG CGGAGGGGGC
151 TCAGCGGGCCG GATCCACGG AAGCGCGCTC GGAGGGGTGG GACCCGGCCG
201 GACCGGAGAT GCGGCCGCCA GCGGGCGGGG CGGCGGCGGC GGCCTCGGAC
251 TTGGGCTCCG CCGCAGTGCT CTGGGCTGTG CACGCCGCGG TGAGGCGGCT
301 GGGCGCCGGG CCAGACGCGG AGGCACAGCT GCGGAGGCTG CAGCTGAGCG
351 CGGACCCTGA GAGGCCTGGG CGCTTCCGGC TGGAGCTGCT GGGCGCGGGA
401 CTTGGGGCGG TTAATTGGA GTGGCCCTG GAGTCAGTTT CCTACACCAT
451 CCGAGGCCCC ACCCAGCAGC AGCTACAGCC TCCACCAGGA GGGCCTGGAA
501 CCCTCAGCCT GCACCTCCTC AACCTCAGG AAGCTCAGCG GTGGGCAGTC
551 CTAGTCCGAG GTGCCACCGT GGAAGGACAG AATGGCAGCA AGAGCAACTC
601 ACCACCAAGC TTGGGCCCCG AAGCATGCCC TGTCTCCCTG CCCAGTCCCC
651 CGGAAGCCTC CACACTCAAG GGCCTCCAC CTGAGGCAGA TCTTCCTAGG
701 AGCCCTGGAA ACTTGACGGA GAGAGAAGAG CTGGCAGGGA GCCTGGCCCG
751 GGCTATTGCA GGTGGAGACG AGAAGGGGGC AGCCCAAGTG GCAGCCGTCC
801 TGGCCCAAGC TCGTGTGGCC CTGAGTGTTC AGCTTCAGGA GGCCTGCTTC
851 CCACCTGGCC CCATCAGGCT GCAGGTCACA CTTGAAGACG CTGCCTCTGC
901 CGCATCCGCC GCGTCTCTG CACACGTTGC CCTGCAGGTC CACCCCACTC
951 GCACCTGTGC AGCTCTCCAG GAGCAGGTCT TCTCAGAGCT CGGTTTCCCG
1001 CCAGCCGTGC AACGCTGGGT CATCGGACGG TGCCTGTGTG TGCCTGAGCG
1051 CAGCCTTGCC TCTTACGGGG TTCGGCAGGA TGGGGACCCT GCTTTCTCTT
1101 ACTTGCTGTC AGCTCCTCGA GAAGCCCCAG CCACAGGACC TAGCCCTCAG
1151 CACCCCAAGA AGATGGACGG GGAACCTGGA CGCTTGTTC CCCCATCATT
1201 GGGGCTACCC CCAGGCCCCC AGCCAGCTGC CTCAGCCTG CCCAGTCCAC
1251 TCCAGCCAG CTGGTCTGTG CCTTCTGCA CCTTCATCAA TGCCCCAGAC
1301 CGCCCTGGCT GTGAGATGTG TAGCACCCAG AGGCCCTGCA CTGGGACCC
1351 CCTTGCTGCA GCTTCCACCT AGCAGCCACC AGAGGTTACA AGGGGAGAGT
1401 GGCCTTCCC TCACAAGTCC GACATCTCCA GGGCCCCACT GAACTCCGGG
1451 GACCTCTACT GACTGCTTGC TGGGACAGTC ACCAGGGTTG GGGGAAGGG
1501 CCACAAAATG AAACCATTA AGACCTTAA GAGCCAAAAA AAAAAAAAAA
1551 AAAAAAAAAA AAAAAAAAAA AAAAAAAG
```

## BLAST Results

No BLAST result

## Medline entries

No Medline entry

## Peptide information for frame 2

ORF from 209 bp to 1369 bp; peptide length: 387



```

1  MAPPAGGAAA  AASDLGSAAV  LLAVHAAVRP  LGAGPDAAEQ  LRRLLQLSADP
51  ERPRGERLEL  LGAGGAVNL  EWQESVSYT  TRGTPHQELQ  PPPGGPTGLS
101  LHFNLPEQEA  RWALVLRGAT  VEPGSGKSN  SPPLAGPEAC  PVSLSPPPEA
151  STLKGPPPEA  DLPRSPGNLT  EREELAGSLA  RAIJGGDEKG  AAQVAAVLAA
201  HRVALSVQLE  EACFPSPGPIR  LVQTLAEDA  AASAASSHVA  ALQVHPHCTV
251  AALQEQQVSE  LGFPFAVQRW  VIGRCLCPE  RLSASYGVRQ  DGDPAFLYLL
301  SAPREAPATG  PSPQHPQKMD  GLGRLFGLPL  LGLPPGPQA  ASSLPSLPQ
351  SWSCPSTFTI  NADPRPGCEM  CSTORPCTWD  PLAAAST

```

No BLASTP hits available

Report for DKFZphtes3 23n19.2

```
SEQ      MAPPAGGAAAAASDLGSAAVLLAVHA AVRPLGAGPD AEALRRQLSADPERPGRFRLEL
SEG      .xxxxxxxxxxxxxxxxxxxxxxxxxxxxx.....
PRD      cccccchhhhhhhhhhhhhhhhhhhhhhhcccccchhhhhhhhhhhccccccccceee
```

```

SEQ      LGAGPGAVNLEWPLESVSYTIRGPTQHELQPPPGGPGTLSLHFLNPQEAQRWAVLVRGAT
SEG
PRD      cccccceeeccccceeeccccccccccccccccccccceeeccccchhhhhheeeccc

SEQ      VEQNGSKSNSPPALGPEACPVSLSPPPEASTLKGPPPEADLPRSPGNLTEREELAGSLA
SEG
PRD      eccccccccccccccccccccccccccccccccccccccccccccccccchhhhhhhhhhhh

SEQ      RAIAGGDEKGAAQVAAVLAQHRVALSVQLQEACFPFGPIRLQVTLEDAASAASAASHAV
SEG
PRD      hhhhhcccchhhhhhhhhhhhhhhhhhhhhccccccccccccceeecccchhhhhhhhhhhhee

SEQ      ALQVHPHCTVAALQEQVSELGFPPAVQRWVIGRCLCPERSLASYGVRQDGPDAFLYLL
SEG
PRD      eeccccchhhhhhhhhhhhhhhccccccccchhhhhhhhhhhhhccccccccccccceeecc

SEQ      SAPREAPATGPSQHPQKMDGELGRLFPPSLGLPGPQPAASSLPSPLQPSWSCPSTFI
SEG
PRD      . . . . . xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx . . . . .
ccccccccchhhhhhhhhhhhhhhccccccccccccccccccccccccccccccccceee

SEQ      NAPDRPGCEMCSTQRPCWDPPLAAST
SEG
PRD      cccccccccccccccccccccceeeccc

```

(No Prosite data available for DKFZphtes3\_23n19.2)

(No Pfam data available for DKFZphtes3\_23n19.2)

similarity to rat protein kinase C-interacting RBCC protein 1

start at Bp 209 matches kozak consensus PyNNatgG  
similarity to of C-terminal part to N-terminus of RBCK1

Sequenced by LMU

Locus: unknown

Insert length: 1579 bp  
Poly A stretch at pos. 1535, polyadenylation signal at pos. 1515

1	CGGAGACCCCT	CGGGCCGTGT	CCATTTGTGG	GCAAAGCCAG	CGGGGAGGCC
51	TGTGGCAGGAT	TGCACCACTC	GGCGCCCGTCC	GACGCCGCAG	GCTCTTGGGGC
101	CGCCCGGAAG	CCAGAGTTCC	GGCGCCGTGT	TCCGACCCGG	CGAGAGGGGCG
151	TGACGGCGCCC	GATCCCAACG	AAGCGCGCTC	GGGGGGGTGG	GACCCCGCGCG
201	CGACGGAGAT	GGCCGCCGCA	CGCGGCGGGG	CGGCGCCGCG	GGCCTTCGACG
251	TTGGGCTCCG	CCGCAGTGCT	CTTGCTGTGT	CACGCCGCGG	TGAGGCGCGG
301	GGGCGCCGGG	CGACAGCCGC	AGGCCACAGT	CGCGAGGCTG	CAGCTTGAGCG
351	CGACCCCTGA	GAGGCCCTGG	CGCTTCGGCG	TGGAGTCTGT	GGGGCGGGGA
401	CTCTGGGGGCG	TTAATTGTGA	TGTGCCCTGT	GAGTCAGTTT	CCTACACCAT
451	CCGAGAGCCC	ACCCAGCAGC	AGTCAACAGA	TCCACAGGA	GGGCTCGGAA
501	CCCTCAGCCT	GCACTTCCCT	AACCTTCAGG	AAGTCACGCA	GTGGGCAGTC
551	CTAGTCCGAG	GTGCCACCGT	GGAAGGACAG	AATGGCAGCA	AGAGCAACTC
601	ACCAACAGCC	TTGGGCCAAG	AGCATAGCCC	TGTCCTCCCT	CCCACTCCCC
651	CGGAAGCTCG	GTGACCAAGA	GGCCCTCCAC	TACGAGGACA	TCCTTCTTAGG
701	AGCCCTGGAA	ACTTGACGAG	GAGAGAAGAG	CTGGACGGGA	GCCCTGGGCCG
751	GGCTATTGTGA	CGTGGAGACG	AGAAAGGGGG	AGCCCAAGT	GACAGCCGCTC
801	TGGCCACGCA	TGCTGTGGCC	CTAGTGTGTC	AGCTTCAGGA	GGCCTTGTTCC
851	CCACCTGGCC	CCATCAGGCT	GCAGGTCACA	CTTGAAGACG	CTGCCCTCTGC
901	CGACTCCGCC	GGCTCCTCTG	CACACGTTGC	CTCCAGGCTG	CACCCCCACT
951	GCACTGTTCG	AGCTCTCAG	GACGAGGTGT	TCTCAGAGCT	CGGTTTCCCG
1001	CAGCGCGTGC	TAGCTTGGTG	CTCCGGAAGC	TGCCCTGTGT	TGCCCTGAGCG
1051	CCAGGCTTGC	ACTTACGGGG	ATCGCGAGGA	TGGGACCCCT	GCTTTCTCTCT
1101	ACTTGCTGTC	AGCTCCTCGA	GAAAGCCGAG	CCACGAGACC	TAGCCCTCAG
1151	CACCCCCAGA	AGATGGACGG	GGAACTTGGA	CGCTTGTTTT	CCCCATCAT
1201	GGGGCTTACC	CAGGCCCCCC	AGCCAGCTGC	CTCCAGCCCT	CCCACTCAC
1251	TCCAGCCGCA	GTTGCTCTGT	CTTCTCTSCA	CTTCTATCAA	TGCCCCAGAC
1301	CGACCTGGCT	GTGAGATGTG	CAGCCAGCCG	AGGCCCTGCA	CTTGGGACAG
1351	CCTTGCTGCA	GCTTCCACCT	AGCAGCCAC	AGAGGTTACA	AGGGGAGAGT
1401	GGCCCTTCCC	TCAACAAGTC	GACATCTCCA	GGGCCCCACT	GAACCTCCGG
1451	GACCTCTACT	GACTGCTTGC	TGGGACAGTC	ACCAAGGTTG	GGGGGAAGGG
1501	CCACAAAATG	AACCACTTAA	AGACCTTTAA	GAGCCAAAAA	AAAAAATAAA
1551	AAAAAATAAA	AAAAAATAAA	AAAAAATAAG		

## BLAST Results

No BLAST result

## Medline entries

No Medline entry

## Peptide information for frame 2

ORF from 209 bp to 1369 bp; peptide length: 387  
 Category: similarity to known protein  
 Classification: Cell signaling/communication

```

1 MAPPAGGAAA AASDLGSAAV LLAVHAAVRP LGAGPDAAEQ LRRQLSADP
51 ERPGFRFRLEL LGAGPGAVNL EWPLESVSYT IRGPTQHELQ PPPGGPGTSL
101 LHFLNPQEAQ RWAVLVRGAT VEGQNGSKSN SPPALGPEAC PVSLPSPPEA
151 STLKGPPEA DLPRSPGNLT EREELAGSLA RAIAGGDEKG AAQVAAVLAQ
201 HRVALSVQLQ EACFPPIR LQVTLEDAAS AASAASSAHV ALQVHPHCTV
251 AALQEQVFSE LGFPPAVQRW VIGRCLCVPE RSLASYGVRQ DGDPAFLYLL
301 SAPREAPATG PSPQHPQKMD GELGRLEFPPS LGLPPGPQPA ASSLPSPLOP
351 SWSCPSCTFI NAPDRPGCEM CSTQRPCTWD PLAAAST

```

## BLASTP hits

No BLASTP hits available

## Alert BLASTP hits for DKFZphtes3\_23n19, frame 2

PIR:JC5983 protein kinase C-interacting RBCC protein 1 - rat, N = 1,  
 Score = 353, P = 2.8e-32

TREMBL:AB011369\_1 product: "RBCK2"; Rattus norvegicus mRNA for RBCK2,  
 complete cds., N = 1, Score = 353, P = 2.8e-32

TREMBL:U67322\_1 gene: "XAP4"; product: "HBV associated factor"; Human  
 HBV associated factor (XAP4) mRNA, complete cds., N = 1, Score = 286, P  
 = 8.5e-25

TREMBLNEW:AF124663\_1 product: "UbcM4 interacting protein 28"; Mus  
 musculus UbcM4 interacting protein 28 mRNA, complete cds., N = 1, Score  
 = 367, P = 9.3e-34

>TREMBLNEW:AF124663\_1 product: "UbcM4 interacting protein 28"; Mus musculus  
 UbcM4 interacting protein 28 mRNA, complete cds.  
 Length = 498

## HSPs:

Score = 367 (55.1 bits), Expect = 9.3e-34, P = 9.3e-34  
 Identities = 95/212 (44%), Positives = 129/212 (60%)

Query: 175 LAGSLARAIAGGDEKGAAQVAAVLAQHRVALSVQLQEACFPPIRLQVTLEDAASAASA 234  
 +A SLARA+AGGDE+ A + A LA+ RV L VQ++ P IRL V++EDA  
 Sbjct: 1 MALSLARAVAGGDEQAAYATWLAEQRVPLRVQVQKPEVSPTQDIRLCVSVEDAYM---- 56

Query: 235 ASSAHVALQVHPHCTVAALQEQVFSELGFPPAVQRWVIGRCLCVPERSLASYGVRQDGDG 294  
 + + L V P TVA+L++ VF + GFPP++Q+VV+G+ L + +L S+G+R++GD  
 Sbjct: 57 -HTVTIWLTVRPDMTVASLKDMVFLDYGFPPSLQQVVGQRLARDQETLHSHGIRNRNGDG 115

Query: 295 AFLYLLSAPREAPATGSPQHPQK-----MDGELG--RLFPPLSG-LPPG-PQPAASSLP 345  
 A+LYLLSA T +PQ Q+ M +LG L S G L P P+P + P  
 Sbjct: 116 AYLYLLSARN----TSLNPQELQRQRLRMLEDLGFKDITLQSRGPLEPVLKPKRTNQEP 171

Query: 346 -----SPLQF--SWSCPSCTFINAPDRPGCEMCSTQRPCTW 379  
 +P P W CP CTFIN P RPGCEMC RP T+  
 Sbjct: 172 GQPDAAPESPVVGWQCPCGCTFINKPTRPGCEMCCRARPETY 212

## Pedant information for DKFZphtes3\_23n19, frame 2

```
[LENGTH]      387
[MW]          39949.29
[pI]          5.53
[HOMOL]       TREMBLNEW:AF124663_1 product: "UbcM4 interacting protein 28"; Mus musculus
UbcM4 interacting protein 28 mRNA, complete cds. 1e-22
[BLOCKS]      BL00578B
[KW]          Alpha_Beta
[KW]          LOW_COMPLEXITY      17.57 %
```

(No Pfam data available for DKFZphtes3\_23n19.2)

DKFZphtes3\_26g22

group: intracellular transport/trafficking

DKFZphtes3\_26g22 encodes a novel 898 amino acid protein with similarity to kinesins.

The novel protein contains a ATP/GTP-binding site motif A (P-loop) and a kinesin motor domain signature. Kinesin is a microtubule-associated force-producing protein that play a role in organelle transport. It is an oligomeric complex composed of two heavy chains and two light chains. The kinesin motor activity is directed toward the microtubule's plus end. The heavy chain contains a large globular N-terminal domain which is responsible for the motor activity of kinesin, which is known to hydrolyze ATP and to bind and move on microtubules. Several proteins involved in chromosome segregation and cell division contain this motor domain, such as drosophila claret segregational protein (ncd), Drosophila kinesin-like protein (nod), human CENP-E and human mitotic kinesin-like protein-1 (MKLP-1). The novel protein is a new kinesin like protpein.

The new protein can find application in modulating chromosome transport in mitosis and meiosis and modulation of cell division.

strong similarity to kinesins

Sequenced by EMBL

Locus: unknown

Insert length: 3032 bp

No poly A stretch found, no polyadenylation signal found

```
1 CTGAAGCGCT GGGAGGCGGA CATTAAAGTG AAGTGGTTGC GGTAACCTGG
51 CCTGGGCGCTG AAGTGAGTGA GAGGCACATG AAGAGAAGTA TTCAAGTATT
101 TATACAGATA GGAATCAAGA TAATCAACAA TGCTGTGCAC TGAGGAAGAC
151 CTGTGCCACC ATATGAAAAGT AGTAGTTCGT GTACGTCGGG AAAACACTAA
201 ACAGAAAAGCA GCTGGATTTC ATAAAGTGGT TCATGTTGTG GATAAACATA
251 TCCTAGTTTT TGATCCCAA CAAGAAGAAG TCAGTTTTTT CCATGGAAAG
301 AAAACTACAA ATCAAAATGT TATAAAGAAA CAAAATAAGG ATCTTAAAT
351 TGTATTTGAT GCTGTTTTTG ATGAAACGTC AACTCAGTCA GAAGTTTTTG
401 AACACACTAC TAAGCCAATT CTTCGTAGTT TTTTGAATGG ATATAATTGC
451 ACAGTACTTG CCTATGGTGC CACTGGTSGT GGGGAAGACCC ACACATGCT
501 AGGATCAGCT GATGAACCTG GAGTGATGTA TCTAACAAATG TTACACCTTT
551 ACAATGCGAT GGATGAGATT AAAGAAGAGA AAATATGTAG TACTGCAGTT
601 TCATATCTGG AGGTATATAA TGAACAGATT CGTGATCTCT TAGTAAATTC
651 AGGGCCACTT GCTGTCCGGG AAGATACCCA AAAAGGGGTG GTCGTTTCATG
701 GACTTACTTT ACACCAGCCC AAATCCTCAG AAGAAATTTT ACATTTATTG
751 GATAATGGAA ACAAAACAG GACACAACAT CCCACTGATA TGAATGCCAC
801 ATCTTCTCGT TCTCATGCTG TTTTCCAAAT TTACTTGCGA CAACAAGACA
851 AAACAGCAAG TATCAATCAA AATGTCCGTA TTGCCAAGAT GTCACCTATT
901 GACCTGGCAG GATCTGAGCG AGCAAGTACT TCCGGTGCTA AGGGGACCCG
951 ATTTGTAGAA GGCACAAATA TTAATAGATC ACTTTTAGCT CTTGGGAATG
1001 TCATCAATGC CTTAGCAGAT TCAAAGAGAA AGAATCAGCA TATCCCTTAC
1051 AGAAATAGTA AGCTTACTCG CTTGTTAAAG GATTCTCTTG GAGGAAACTG
1101 TCAAACATATA ATGATAGCTG CTGTTAGTCC TTCTCTGTA TTCTACGATG
1151 ACACATATAA CACTCTTAAG TATGCTAACC GGGCAAAAGG CATTAAATCT
1201 TCTTTGAAGA GCAATGTTCT TAATGTCAAT AATCATATAA CTCAATATGT
1251 AAAGATCTGT AATGAGCAGA AGGCAGAGAT TTTATTGTTA AAAGAAAAAC
1301 TAAAGCCTTA TGAAGAACAG AAAGCCTTCA CTAATGAAAA TGACCAAGCA
1351 AAGTTAATGA TTTCAAACCC TCAGGAAAAA GAAATCGAAA GGTTTCAAGA
1401 AATCCTGAAC TGCTTGTTC AGAATCGAGA AGAAATTAGA CAAGAATATC
1451 TGAAGTTGGA AATGTTACTT AAAGAAAATG AACTTAAATC ATTCTACCAA
1501 CAACAGTGCC ATAAACAAAT AGAAATGATG TGTTCGTAAG ACAAAGTAGA
1551 AAAGGCCACT GGAACACGAG ATCATAGACT TGCAATGTTG AAAACTCGTC
1601 GCTCCTACCT GGAGAAAAGG AGGGAGGAGG AATTGAAGCA ATTTGATGAG
1651 AATACTAATT GGCTCCATCG TGTCGAAAAA GAAATGGGAC TCTTAAGTCA
1701 AAACGGTCAT ATTCCAAAGG AACTCAAGAA AGATCTTCAT TGTCACCATT
1751 TGCACCTCCA GAACAAAGAT TTGAAAGCAC AAATTAGACA TATGATGGAT
1801 CTAGCTTGTC TTCAGGAACA GCAACACAGG CAGACTGAAG CAGTATTGAA
1851 TGCTTTACTT CCAACCTTAA GAAAACAATA TTGCACATTA AAAGAAGCCG
1901 GCCTGTGCAA TGCTGCTTTT GAATCTGACT TCAAAGAGAT CGAACATTTG
1951 GTAGAGAGGA AAAAAGTGGT AGTTTGGGCT GACCAAACTG CCGAACACC
2001 AAAGCAAAAC GATCTACCAG GGATTTCTGT TCTTATGACC TTTCCACAAC
2051 TTGGACAGT TCAGCCTATT CTTGTTGCT CATCTTCAGG TGGAACTAAT
2101 CTGTTTAAAG TTCCTACAGA AAAAAGAACT CGGAGAAAAC TAATGCCATC
2151 TCCCTTGAAA GGACAGCATA CTCTAAAGTC TCCACCATCT CAAAGTGTGC
2201 AGCTCAATGA TTCTCTTAGC AAAGAACTTC AGCCTATTGT ATATACACCA
2251 GAACAGCTGA GAAAAGCTTT TCAAAATCCG TCTACAGTAA CCTTAATGAA
2301 ACCATCATCA TTTACTACAA GTTTTCAGGC TATCAGCTCA AACATAAACA
2351 GTGATAATTG TCTGAAAAATG TTGTTGTAAG TAGCTATCCC TCATAATAGA
```

```

2401 AGAAAAGAAT GTGGACAGGA GGAATTGGAC TCTACATTTA CTATATGTGA
2451 AGACATCAAG AGCTCGAAGT GTAAATTACC CGAACAAGAA TCACTACCAA
2501 ATGATAACAA AGACATTTTA CAACGGCTTG ATCCTTCTTC ATTCTCAACT
2551 AAGCATTCTA TGCCTGTACC AAGCATGGTG CCATCCTACA TGGCAATGAC
2601 TACTGCTGCC AAAAGGAAAC GGAAATTAAC AAGTTCTACA TCAAAACAGTT
2651 CGTTAACTGC AGACGTAAAT TCTGGATTTG CCAAACGTGT TCGACAAGAT
2701 AATTCAAGTG AGAAGCACTT ACAAGAAAAC AAACCAACAA TGGAACATAA
2751 AAGAAACATC TGTAAATATA ATCCAAGCAT GGTAGAAAAA TTTGGAAGAA
2801 ATATTTCAAA AGGAAATCTA AGATAAATCA CTTCAAAACC AAGCAAAATG
2851 AAGTTGATCA AATCTGCTTT TCAAAGTTTA TCAATACCTT TTTCAAAATA
2901 TATTTAAAT CTTTGAAAGA AGACCCATCT TAAAGCTAAG TTTACCAAG
2951 TACTTTCAGC AAGCAGAAAA ATGAAACTCT TTGTTTCTT CTTTGTGTT
3001 CTAATAAAAT AAAATTTCAA AAGAAAAAAA AA

```

## BLAST Results

-----

No BLAST result

## Medline entries

-----

No Medline entry

## Peptide information for frame 1

-----

ORF from 130 bp to 2823 bp; peptide length: 898  
 Category: strong similarity to known protein  
 Classification: Cell structure/motility  
 Prosite motifs: ATP\_GTP\_A (113-121)  
 KINESIN\_MOTOR\_DOMAIN1 (252-264)

```

1 MSVTEEDLCH HMKVVVRVRP ENTKEKLAGF HKVVHVVDKH ILVFDPKQEE
51 VSFFHGKKT NQNVIKKQNK DLKFVFDVAF DETSTQSEVF EHTTKPILRS
101 FLNGYNCTVL AYGATGAGKT HTMLGSADEP GVMYLTMLHL YKCMDEIKKE
151 KICSTAVSYL EVYNEQIRDL LVNSGPLAVR EDTQKGVVVH GLTLHQPKSS
201 EEILHLLDNG NKNRTQHPTD MNATSSRSHA VFQIYLRQDD KTASINQNV
251 IAKMSLIDLA GSERASTSGA KGTRFVEGTN INRSLLALGN VINALADSKR
301 KNQHIPPYRN KLTRLKDSL GGNCQTIMIA AVSPSSVFYD DTYNLTLYAN
351 RAKDIKSSLK SNVLNVNNHI TQYVKICNEQ KAEILLKEK LKAYEEQKAF
401 TNENDQAKLM ISNPQKEIE RFQEILNCLF QNREEIRQEQ LKLEMLLKEN
451 ELKSFYQQQC HKQIEMMCSE DKVEKATGKR DHRRLAMKTR RSYLEKRRKE
501 ELKQFEDENT WLHRVEKEMG LLSQNGHIPK ELKLDLHCHH LHLQNKDLKA
551 QIRHMDLAC LQEQQHQTE AVLNALLPTL RKQYCTLKEA GLSNAAFESD
601 FKEIEHLVER KKVVVWADQT AEQPKQNDLP GISVLMTFPQ LGPVQPIPC
651 SSSGGTNLVK IPTEKTRRK LMPSPKQGH TLKSPSPQSV QLNDSLKSEL
701 QPIVYTPEDC RKAFQNPSTV TLMKPSSFTT SFQAISNNIN SDNCLKMLCE
751 VAIPHNRKE CGQEDLDSTF TICEKIKSSK CKLPEQESLP NDNKDILQRL
801 DPSSFSTKHS MPVPSMVPST MAMTTAAKRK RKLTSSTNS SLTADVNSGF
851 AKRVQDNSS EKHLQENKPT MEHKNRICKI NPSMVRKEGR NISKGNLR

```

## BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphtes3\_26g22, frame 1

SWISSPROT:YB3D\_SCHPO PUTATIVE KINESIN-LIKE PROTEIN C2F12.13., N = 3,  
 Score = 874, P = 9e-93

TREMBL:DMU89264\_1 product: "kinesin like protein 67a"; Drosophila  
 melanogaster kinesin like protein 67a mRNA, complete cds., N = 1, Score  
 = 880, P = 4.2e-88

TREMBL:SPBC649\_1 gene: "SPBC649.01c"; product: "putative kinesin-like  
 protein"; S.pombe chromosome II cosmid c649., N = 3, Score = 814, P =  
 9.8e-86

PIR:S64238 kinesin-related protein KIP3 - yeast (Saccharomyces  
 cerevisiae), N = 2, Score = 802, P = 2.5e-83

>TREMBL:DMU89264\_1 product: "kinesin like protein 67a"; Drosophila

melanogaster kinesin like protein 67a mRNA, complete cds.  
Length = 814

## HSPs:

Score = 880 (132.0 bits), Expect = 4.2e-88, P = 4.2e-88  
Identities = 181/345 (52%), Positives = 238/345 (68%)

```
Query: 11 HMKVVRVRPENTKEAAGFHKVHVVDKHLVDPKQEEVSFF-HGKKTNNQNVIKKQN 69
      ++KV VRVRP N +E      ++ V+D+ L+FDP +E+ FF G K +++ K+ N
Sbjct: 8 NIKVAVRVRPYNVRELEQKQRSIIKVMDSALLFDPDEDEDEFFFGAKQPYRDITKRMN 67

Query: 70 KDLKFVFDVAFDETSTQSEVFETHTKPIILRSFLNGYNCTVLAYGATGAGKTHMLGSADE 129
      K L FD VFD ++ ++FE T P++ + LNGYNC+V YGATGAGKT TMLGS
Sbjct: 68 KKLTMEFDRVFDIDNSNQDLFEECTAPLVDAVLNGYNCVVFYVYATGAGKTFTMLGSEAH 127

Query: 130 PGVMYLTMLHLYKCMDEIKEEKICSTAVSYLEVYNEQIRDLLVNSGPLAVREDTQKGVVV 189
      PG+ YLTM L+ + + + VSYLEVYNE + +LL SGPL +RED GVVV
Sbjct: 128 PGLTYLTMQDLFDKIQAQSDVRKFDVGSYLEVYNEHVMNLLTKSGFLKLRDNN-GVVV 186

Query: 190 HGLTLHQPKSSEIHLHLDNGNKNRTQHPTDMNATSSRSHAVFQIYLRQQDKTASINQNV 249
      GL L S+EE+L +L GN +RTQHPTD NA SSRSHA+Q+++R ++ + V
Sbjct: 187 SGLCLTPIYSAEELLRLMLGNSHRTQHPTDANAESSRSHAIFQVHIRITERKTDTKRTV 246

Query: 250 RIAKMSLIDLAGSERASTSGAGKTRFVEGTNINRSLLAGNVINALADSKRKNQHIPYRN 309
      K+S+IDLAGSERA+++ G RF EG +IN+SLLALGN IN LAD + HIPYR+
Sbjct: 247 ---KLSMIDLAGSERAASTKGIGVRFKEGASINKSLLALGNCINKLADGLK---HIPYRD 300

Query: 310 SKLTRLLKDSLGGNCQTIMIAAVSPSSVFYDDTYNTLKYANRAKDI 355
      S LTR+LKDSLGGNC+T+M+A VS SS+ Y+DTYNTLKYA+RAK I
Sbjct: 301 SNLTRILKDSLGGNCRTLMVANVSMSSLTEDTYNTLKYASRAKKI 346
```

Pedant information for DKFZphtes3\_26g22, frame 1

## Report for DKFZphtes3\_26g22.1

```
[LENGTH] 898
[MW] 102281.63
[pI] 9.09
[OMOL] SWISSPROT:YB3D SCHPO PUTATIVE KINESIN-LIKE PROTEIN C2F12.13. 3e-97
[FUNCAT] 30.04 organization of cytoskeleton [S. cerevisiae, YGL216w] 2e-88
[FUNCAT] 03.22 cell cycle control and mitosis [S. cerevisiae, YGL216w] 2e-88
[FUNCAT] 08.22 cytoskeleton-dependent transport [S. cerevisiae, YGL216w] 2e-88
[FUNCAT] 30.10 nuclear organization [S. cerevisiae, YGL216w] 2e-88
[FUNCAT] 09.10 nuclear biogenesis [S. cerevisiae, YPR141c] 5e-42
[FUNCAT] 06.10 assembly of protein complexes [S. cerevisiae, YPR141c] 5e-42
[FUNCAT] 03.13 meiosis [S. cerevisiae, YPR141c] 5e-42
[FUNCAT] 11.01 stress response [S. cerevisiae, YPR141c] 5e-42
[FUNCAT] 03.07 pheromone response, mating-type determination, sex-specific proteins
      [S. cerevisiae, YPR141c] 5e-42
[FUNCAT] 30.05 organization of centrosome [S. cerevisiae, YPR141c] 5e-42
[FUNCAT] 03.04 budding, cell polarity and filament formation [S. cerevisiae, YKL079w]
4e-28
[BLOCKS] BL00411H
[BLOCKS] BL00411G
[BLOCKS] BL00411F
[BLOCKS] BL00411E Kinesin motor domain proteins
[BLOCKS] BL00411C Kinesin motor domain proteins
[BLOCKS] BL00411B Kinesin motor domain proteins
[BLOCKS] BL00411A Kinesin motor domain proteins
[SCOP] d2kin.1 3.29.1.5.3 Kinesin [Rat (Rattus norvegicus) 1e-117
[SCOP] d3kar_ 3.29.1.5.4 Kinesin [Baker's yeast (Saccharomyces 1e-112
[PIRKW] nucleus 6e-87
[PIRKW] heterodimer 4e-68
[PIRKW] DNA binding 9e-60
[PIRKW] heterotetramer 2e-54
[PIRKW] mitosis 9e-60
[PIRKW] microtubule binding 4e-68
[PIRKW] ATP 6e-87
[PIRKW] phosphoprotein 5e-59
[PIRKW] heterotrimer 4e-68
[PIRKW] purine nucleotide binding 1e-26
[PIRKW] P-loop 6e-87
[PIRKW] coiled coil 4e-68
[PIRKW] heptad repeat 3e-62
[PIRKW] methylated amino acid 2e-54
[PIRKW] hydrolase 2e-54
[PIRKW] GTP binding 1e-60
```

```

[PIRKH]      cell division 5e-57
[SUPFAM]     kinesin-related protein KIP1 3e-50
[SUPFAM]     kinesin-related protein CIN8 7e-33
[SUPFAM]     kinesin heavy chain 2e-54
[SUPFAM]     suppressor protein SMY1 1e-26
[SUPFAM]     kinesin-related protein KIF3 4e-68
[SUPFAM]     kinesin-related protein KIF2 1e-46
[SUPFAM]     kinesin-related protein unc-104 7e-60
[SUPFAM]     unassigned kinesin-related proteins 6e-87
[SUPFAM]     centromere protein E 3e-54
[SUPFAM]     kinesin-related protein KLP61F 5e-57
[SUPFAM]     kinesin-related protein MKLP-1 2e-28
[SUPFAM]     pleckstrin repeat homology 7e-60
[SUPFAM]     kinesin-related protein KIF1B 4e-61
[SUPFAM]     kinesin motor domain homology 6e-87
[SUPFAM]     kinesin-related protein KLP4 1e-43
[SUPFAM]     kinesin-related protein nodA 1e-30
[SUPFAM]     kinesin-related protein Eg5 5e-59
[PROSITE]    ATP_GTP_A 1
[PROSITE]    KINESIN_MOTOR_DOMAIN1 1
[PFAM]       Kinesin motor domain
[KW]         Irregular
[KW]         3D
[KW]         LOW_COMPLEXITY 8.57 %

```

```

SEQ      MSVTEEDLCHMMKVVRVRPENTKEKAAGFHKKVHVVDKHLVDFDPKQEEVSFFHGKKT
SEG      .....
3kar-    .....TBEEE

SEQ      NQNVIKKQNKDLKFVDAVFDETSTQCEVFEHTTKPILRSFLNGYNTVLAYGATGAGKT
SEG      .....
3kar-    EEEEETTTTTEEEEEETETTTTCHHHHHHHHHH-HHHGGGGCCCEEEECTTTTCHH

SEQ      HTMLGSADEPGVMYLTMLHLYKCMDEIKEEKICSTAVSYLEVYNEQIRDLLVNSGPLAVR
SEG      .....
3kar-    HHHHTTTT--THHHHHHHHHHHHHHHGGGCEEEEEEEEEETTEEEETT-TCCCEEE

SEQ      EDTQKGVVHGLTLHQPKSSEIHLHLLDNGNKNRTQHPDMNATSSRSHAVFQIYLRQOD
SEG      .....
3kar-    EETTTTEEEETTCCCEEECCGGHHHHHHHHHHHCCCTTTTCHHHHHHCEEEEEEEEEEE

SEQ      KTASINQNVRIAKMSLIDLAGSERASTSGAKGTRFVEGTNINRSLLAGNVINALADSKR
SEG      .....
3kar-    TTTTCEE---EEEEEEECCECCCCCCC---HHHHHHHHHHHHHHHHHHHHHHHTTTT

SEQ      KNQHIPPYRNSKLTRLKDSLGGNCQTIMIAAVSPSSVFYDDTYNTLKYANRAKDIKSSLK
SEG      .....
3kar-    TTTCTTTTTHHHHHHHGGGCTTTTEEEEEEECCGGHHHHHHHHHHHH.....

SEQ      SNVLNVNHHITQYVKICNEQKAEILLKEKLKAYEEQKFTNENDQAKLMISNPQEKEIE
SEG      xxxxxxxx.....
3kar-    .....

SEQ      RFQEILNCLFQNRREEIRQEYLKLEMLLKENELKSFYQQOCHKQIEMMCSEDKVEKATGKR
SEG      .....
3kar-    .....

SEQ      DHRLAMLKTRRSYLEKRREELKQFDENTNWLHRVEKEMGLLSQNGHIPKELKKDLHCHH
SEG      .....
3kar-    .....

SEQ      LHLQNKDLKAQIRHMDLACLQEQHRQTEAVLNALLPTLRKQYCTLKEAGLSNAAFESD
SEG      xxx.....
3kar-    .....

SEQ      FKEIEHLVERKKVVVWADQTAEQPKQNDLPGISVLMTFPQLGPVQPIPCSSSSGGTNLVK
SEG      .....
3kar-    .....

SEQ      IPTEKRRRLKMPSPKLGQHTLKSPPSQSVQLNDSLSKELQPIVYTPEDCRKAFQNPSTV
SEG      .....
3kar-    .....

SEQ      TLMKPSSFTTSFQAISNINSNCLKMLCEVAIPHNRKKECGQEDLDSTFTICEDIKSSK
SEG      .....
3kar-    .....

SEQ      CKLPQESLPNDNKDILQRLDPSSFSTKHSMPVPSMVPSYAMTTAAKRRKLTSTSTNS
SEG      .....
3kar-    .....

```



SEQ SLTADVNSGFAKRVQRQDNSSEKHLQENKPTMEHKRNICKINPSMVRKFGRNISKGNLR  
 SEG xxx.....  
 3kar- .....

## Prosites for DKFZphtes3\_26g22.1

PS00017 113->121 ATP\_GTP\_A PDOC00017  
 PS00411 252->264 KINESIN\_MOTOR\_DOMAIN1 PDOC00343

## Pfam for DKFZphtes3\_26g22.1

HMM_NAME	Kinesin motor domain		
HMM	*RCRPLNeReindgcscvVQWpPwtGyktvhngheds.....		
Query	17	RVRPENTKEKAAGFHKKVVHVVD-KHILVFDPKQEEVSFFHGKKTNQNV	64
HMM	.....phksFtFDHVFwWncTQedVYdtvAHPIVDDcFhGYNCTIFAYGQ		
Query	65	IKKQNKDLKFVFDVDETSTQSEVFEHTTKPILRSFLNGYNCTVLAYGA	114
HMM	TGSGKTYTMMGpggehPDHmGIIPRCCHDIFdrIdkfgekDhdFWHvKCS		
Query	115	TGAGKTHMLG---SADEPGVMYLTMLHLYKCMDEIK-EEKIC-STAVS	158
HMM	YMEIYNeeIYDLLCPnPgHmKpLnIHEHPNMgpyVqGCTefHvCSYeDac		
Query	159	YLEVYNEQIRDLLV-N---SGPLAVREDTQKGVVHGLTLHQPKSSEEIL	204
HMM	hWIWqGnknRHVAaTnMNdhSSRShtIFTIHVeQrHk..qcdehvcHskM		
Query	205	HLLDNGNKNRTQHPTDMNATSSRSHAVFYILRQDKTASINQNVRIAKM	254
HMM	NLVDLAGSERvnrTGAEGQRlKEGcNINqSLttLGnVInaLaDgqTKYmY		
Query	255	SLIDLAGSERASTSGAGKTRFVEGTNINRSLALGNVINALADSK-----	299
HMM	gghgHIPYRDSKLTWLLQDSLGGNcKtcmIACIWPadWNYEETLSTLRYA		
Query	300	RKNQHIPPYRNSKLTLLKDSLGGNCQTIMIAAVSPSSVFYDDTYNTLKYA	349
HMM	dRAKnIkNkPQINEDPcamalWRrYheQIqdmKKhqL*		
Query	350	NRADIKSSLSKSNVLNVN-NHITQYVKICNEQKAEI	384

DKFZphtes3\_27d1

group: metabolism

DKFZphtes3\_27d1 encodes a novel 712 amino acid protein similar to ubiquitin-specific proteases (EC 3.1.2.15).

The novel protein contains both, a ubiquitin carboxyl-terminal hydrolases family 2 signature 1 and signature 2. Pfam predicts a new member of the ubiquitin carboxyl-terminal hydrolases family 2. The ubiquitin system is responsible for the turn over of proteins. Ubiquitin carboxyl-terminal hydrolases (EC 3.1.2.15) (UCH) (deubiquitinating enzymes) are thiol proteases that recognize and hydrolyze the peptide bond at the C-terminal glycine of ubiquitin. These enzymes are involved in the processing of poly-ubiquitin precursors as well as that of ubiquitinated proteins. The novel protein is a new member of the ubiquitin carboxyl-terminal hydrolases family 2, represented by proteins such as yeast UBP1-16, human tre-2, human isopeptidase T and others.

The novel protein can find application in modulation of ubiquitin- and protein metabolism in cells.

similarity to ubiquitin-specific proteases

complete cDNA, complete cds, 4 EST hits

Sequenced by GBF

Locus: unknown

Insert length: 2871 bp

Poly A stretch at pos. 2836, no polyadenylation signal found

```

1  CCAAACCTGA AAGAGGTGA TTTGTAATGA TTTGCAGGGG GGCACCTGGAG
51  GCAGCGGCCA GGACTTTTCA CTTAGGAGAT CAGCATTGTC CCTGATGGAA
101 ACTGGGCGGAT CCTGCAGGGA CTGACCTCTG AGTTATCCAA AGGCCGACCT
151 GGGGAAAGAC TGATTTTGGAG GTTTTAAATAG TTTTCAGATG CTTCAAAGTGT
201 TGTGAACAGA GACTTGTGTT GATTATGCAT TTCTCAGCTA GACTAAATAA
251 ATGCTAGCAA TGGATACGTG CAAACATGTT GGGCAGCTGC AGCTTGCTCA
301 AGACCATTCC AGCCTCAACC CTCAGAAATG GCACTGTGTG GACTGCAACA
351 CGACCGAGTC CATTGCGGCT TGCCTTAGCT GCTCCCATGT TGCCTGTGGA
401 AGATATATTG AAGAGCATGC ACTCAAGCAC TTCAAGAAA GCAGTCATCC
451 TGTTCGATTG GAGGTGAATG AGATGTACGT TTTTGTGTTAC CTTTGTGATG
501 ATTATGTTCT GAATGATAAC GCAACTGGAG ACCTGAAGTT ACTACGACGT
551 ACATTAAGTG CCATCAAAAG TCAAAATTAT CACTGCACAA CTCGTAGTGG
601 GAGGTTTTTA CGGTCCATGG GTACAGGTGA TGATTCTTAT TTCTTACATG
651 ACGGTGCCCA ATCTCTGCTT CAAAGTGAAG ATCAACTGTA TACTGCTCTT
701 TGGCACAGGA GAAGGATACT AATGGGTAAA ATCTTCGAA CATGGTTTGA
751 ACAATCACCC ATTGGAAGAA AAAAGCAAGA AGAACCATTT CAGGAGAAAA
801 TAGTAGTAAA AAGAGAAGTA AAGAAAAGAC GGCAGGAATT GGAGTATCAA
851 GTTAAAGCAG AATTGGAAAG TATGCCTCCA AGAAGAGATT TACGTTTACA
901 AGGGCTCGCT CAGTCGACCA TAATAGAAAT AGTTTCTGTT CAGGTGCCAG
951 CACAAACGCC AGCATCACCA GCAAAAGATA AAGTACTCTC TACCTCAGAA
1001 AATGAAATAT CTCAAAAGT CAGTGACTCC TCAGTAAAC GAAAGCCAAT
1051 AGTAACTCCT GGTGTAACAG GATTGAGAAA TTTGGGAAAT ACTTGCTATA
1101 TGAATTCCTG TCTTCAGGTG TTGAGTCATT TACTTATTTT TCGACAATGT
1151 TTTTAAAGC TTGATCTGAA CCAATGGCTG GCTATGACTG CTAGCGAGAA
1201 GACAAGATCT TGTAAGCATC CACCAGTCAC AGATACAGTA GTATATCAAA
1251 TGAATGAATG TCAGGAAAAA GATACAGGTT TTGTTTGTCT CAGACAATCA
1301 AGTCTGTCTA CAGGACTAAG TGGTGGAGCA TCAAAAGGTA GAAAGATGGA
1351 ACTTATTCAG CCAAAGGAGC CAACTTCACA GTACATTCTT CTTTGTCTATG
1401 AATTGCATAC TTTGTTCCAA GTCATGTGGT CTGGAAAGTG GGCCTTGGTC
1451 TCACCATTTG CTATGCTACA CTCAGTGTGG AGACTCATTC CTGCCCTTTCG
1501 TGGTTACGCC CAACAAGACG CTCAGGAATT TCTTTGTGAA CTTTGTAGATA
1551 AAATACAACG TGAATTAGAG ACAACTGGTA CCAGTTTACC AGCTCTTATC
1601 CCCACTTCTC AAAGGAAACT CATCAACAA GTTCTGAATG TTGTAATATA
1651 CATTTTTCAT GGACAACCTC TTAGTCAGGT TACATGCTCT GCATGTGACA
1701 ACAAAATCAA TACCATAGAA CCTTCTGGG ACTTGTCTAT GGAGTTTCCA
1751 GAAAGGTATC AATGCAGTGG AAAAGATATT GCTTCCCAGC CATGTCGTGT
1801 TACTGAAATG TTGGCCAAAT TTACAGAAAC TGAAGCTTTA GAAGGAAAAA
1851 TCTAGGTATG TGACCAAGTGT AACTCAAAGC GTAGAAGGTT TTCCTCCAAA
1901 CCAGTTGTAC TCACAGAAGC CCAGAAACAA CTTATGATAT GCCACCTACC
1951 TCAGGTCTCT AGACTGCACC TCAAACGATT CAGGTGTGTA GGACGTAATA
2001 ACCGAGAGAA GATTGGTGTG CATGTTGGCT TTGAGGAAAT CTTAAACATG
2051 GAGCCCTATT GCTGCAGGGA GACCTGAAA TCCTCAGAC CAGAAATGCTT
2101 TATCTATGAC TTGTCCGCGG TGGTGTATGCA CCATGGGAAA GGATTGGGCT
2151 CAGGGCACTA CACTGCCTAC TGCTATAATT CTGAAGGAGG GTTCTGGGTA
2201 CACTGCAATG ATTCCAAACT AAGCATGTGC ACTATGGATG AAGTATGCAC
2251 GGCTCAAGCT TATATCTTGT TTTATACCCA ACGAGTTACT GAGAATGGAC

```

```

2301 ATTCTAAACT TTTGCCTCCA GAGCTCCTGT TGGGGAGCCA ACATCCCAAT
2351 GAAGACGCTG ATACCTCGTC TAATGAAATC CTTAGCTGAT CCAAAGACAA
2401 TGGGGTTTTC TTCCTGTGAT TTATATATAT ACTTTTAAAG AGACTGATGT
2451 ACCATTTTAA ACTTCATTTT TTCTTGTGAA TCAGTGTATA CTACATTTAT
2501 ACATTTTATA TCTAACAATT TTTTCTTTT ACAGAGTATA AATGTATATA
2551 TCAACTGAAG GTAACACTTT TTTTCATATT TGGAGTTTAA AACTTTTGGT
2601 GTTTACCTCA GACTGATGTT ACCTCTTTTA TATTTTATG TCTTAATTGG
2651 CTCGGATGAT GAACCTGTGC AATCTTCTAC CAACAAAGTT CAAGTGGCAT
2701 CATTTTATAT ACATGTATCT TTTTCAGGTA TTTTCTATAC AAATTCTTAA
2751 TAGATGGAAA ATTAGACTCT AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA
2801 AAAAAAAAAA AAAAAAAAAA AAGGGGCGGC CGCTCTAAAA AAAAAAAAAA
2851 AAAAAAAAAA AAAAAAAAAA G

```

## BLAST Results

No BLAST result

## Medline entries

98072201:  
Regulation of ubiquitin-dependent processes by deubiquitinating enzymes.

98431658:  
The ubiquitin system.

## Peptide information for frame 2

ORF from 251 bp to 2386 bp; peptide length: 712  
Category: similarity to known protein  
Prosite motifs: UCH\_2\_1 (274-290)  
UCH\_2\_2 (619-638)  
UCH\_2\_2 (619-638)

```

1 MLAMDTCKHV GQLQLAQDHS SLNPQKWHCV DCNTTESIWA CLSCSHVACG
51 RYIEEHALKH FQESSHPVAL EVNEMYVFCY LCDDYVLNDN ATGDLKLLRR
101 TLSAIKSONY HCTTRSGRFL RSMGTGDDSY FLHDGAQSLL QSEDQLYTLA
151 WHRRRILMGK IFRTWFEQSP IGRKKQEEPF QEKIVVKREV KKRRQELEYP
201 VKAELESMPK RKSRLRLQGLA QSTIIIEIVSV QVPAQTPASP AKDKVLSTSE
251 NEISQKVSDES SVKRRPIVTP GVTGLRNLGN TCYMNVSVLQV LSHLLIFRQC
301 FLKLDLNLQWL AMTASEKTRS CKHPPVTDTV VYQMNCEQEK DTGFVCSRQS
351 SLSSGLSGGA SKGRKMELIQ PKEPTSQYIS LCHLHTLFQ VMWSGKWALV
401 SPFAMLHSVW RLIPAFRGYA QQDAQEFLCE LLDKIQRELE TTGTSPLPALI
451 PTFQRKLKIQ VLNVVNNIFH GQLLSQVTCF ACDNKSNTIE PFWDLSLEFP
501 ERYQCSGKDI ASQPCLVTEM LAKFTETEAL EGKIYVCDQC NSKRRRFSSK
551 PVVLTEAQKQ LMICHLPOVL RLHLKRRFRWS GRNNREKIGV HVGFEIILNM
601 EPYCCRETLK SLRPECFIYD LSAVVMHGHK GFGSGHYTAY CYNSEGGFWV
651 HCNDSKLSMC TMDEVCKAQA YILFYTQRTV ENHGSKLLPP ELLLSQHPN
701 EDADTSSNEI LS

```

## BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphtes3\_27d1, frame 2

PIR:S57591 hypothetical protein YMR223w - yeast (Saccharomyces cerevisiae), N = 4, Score = 218, P = 8.4e-38

SWISSPROT:UBPB\_HUMAN UBIQUITIN CARBOXYL-TERMINAL HYDROLASE 11 (EC 3.1.2.15) (UBIQUITIN THIOL ESTERASE 11) (UBIQUITIN-SPECIFIC PROCESSING PROTEASE 13) (DEUBIQUITINATING ENZYME 11) (K1AA0055), N = 2, Score = 300, P = 9.3e-31

TREMBL:AF079565\_1 gene: "Ubp41"; product: "ubiquitin-specific protease UBP41"; Mus musculus ubiquitin-specific protease UBP41 (Ubp41) mRNA, complete cds., N = 3, Score = 187, P = 8.7e-30

PIR:I58376 hypothetical protein unip - mouse, N = 3, Score = 214, P = 1.2e-28

>SWISSPROT:UBPB\_HUMAN UBIQUITIN CARBOXYL-TERMINAL HYDROLASE 11 (EC 3.1.2.15)  
 (UBIQUITIN THIOLESTERASE 11) (UBIQUITIN-SPECIFIC PROCESSING PROTEASE 13)  
 (DEUBIQUITINATING ENZYME 11) (KIAA0055).  
 Length = 1,118

## HSPs:

Score = 300 (45.0 bits), Expect = 9.3e-31, Sum P(2) = 9.3e-31  
 Identities = 95/301 (31%), Positives = 149/301 (49%)

Query: 381 LCHEHLTLFQVMWSGKVALVSPFAMLSVWRLIPAFRGYAQQDAQEFLCELDDKIQREL- 439  
 + E + + +W+G++ +SP ++ ++ F GY+QD+QE L L+D + +L  
 Sbjct: 826 VAEFGIIMKALWTGQYRYISPKDFKITIGKINDQFAGYSQDSQELLFLMDGLHEDLN 885

Query: 440 -----ETTGTSLPALIPTSQRKLIKQVLN--VVNNIFHGQLLSQVTCCLADNKSNT 488  
 E L + LN ++ +F GQ S V CL C KS T  
 Sbjct: 886 KADNRKRYKEENNDHLDDFKAAEHAWQKHQNLNESIIVALFQGFQKSTVQCLTCHKKSRT 945

Query: 489 IEPFWDLSLEFFPERYQCSGKDIASQCLVTEMLAKFTETELEGKIYVCDQCNSKRRRFS 548  
 E F LSL +C+ +D CL + +K E + + + C C ++R  
 Sbjct: 946 FEAFMYLSLPLASTSKCTLQD-----CL--RLFSK--EEKLTDNNRFYCSHCRARR----- 992

Query: 549 SKPVVLTEAQKQLMICHLPOVLRHLKRFWRSGRNNREKIGVHVGE-EILNMEPYCC-- 605  
 ++ K++ I LP VL +HLKRF + GR ++K+ V F E L++ Y  
 Sbjct: 993 -----DSLKKIEIWKLPVLLVHLKRFSDYDGRW-KQKLQTSVDPLENLDLSQYVIGP 1044

Query: 606 RETLKSRLPECFIYDLSAVVMHHGKGFSGHYTAYCYNSEGGFWVHCNDSKLSMCTMDEV 665  
 + LK Y+L +V H+G G GHYTAYC N+ W +D ++S ++ V  
 Sbjct: 1045 KNNLKK-----YNLFVSNNHYG-GLDGGHYTAYCKNAARQWFKFDDHEVSDISVSSV 1096

Query: 666 CKAQAYILFYTQ---RVTE 681  
 + AYILFYT RVT+  
 Sbjct: 1097 KSSAAYILFYTSLGPRVTD 1115

Score = 126 (18.9 bits), Expect = 9.3e-31, Sum P(2) = 9.3e-31  
 Identities = 41/116 (35%), Positives = 63/116 (54%)

Query: 200 QVKAEELESMPPR--KSLRLQGLAQSTIIEIVSVQVPAQTPASPAKDKVLSTSENEISQKV 257  
 Q+ AE + P + +S + Q+ I+ + P TP ++K + EIS ++  
 Sbjct: 701 QIPAERDREPSKLKRSYSSPDITQA--IQEEKKRPTVPTVNRENKPTCYPKAEIS-RL 757

Query: 258 SDSSVKKR-RPIVT---PGVTGLRNLGNTCYMNSVLQVLS---HLLIF--RQCFLKLDLNQ 308  
 S S ++ P+ P +TGLRNLGNTCYMNS+LQ L HL + R C+ D+N+  
 Sbjct: 758 SASQIRNLNPVFGSGPALTGLRNLGNTCYMNSILQCLCNAPHLADYFNRNRYQD-DINR 816

Score = 50 (7.5 bits), Expect = 8.3e-23, Sum P(2) = 8.3e-23  
 Identities = 29/106 (27%), Positives = 51/106 (48%)

Query: 173 RKKQEEPFQEKIVVKREVKKRRQELEYQVKAEELESMPPRKSLRLQGLAQSTIIEIVSVQV 232  
 + KQE+ +E+ +++ K R++E E + K + E+ + Q A+ + + S Q  
 Sbjct: 475 KNKQEKELRERQEQEKEKLRKEEQEKAKKKQEA-EENEITEKQKQAKEEMKKESEQA 533

Query: 233 PAQ---TPASPAKD---KVLSTSENEIS--QKVSDSSVKKRPIVTPGV 272  
 + T A K+ K S SE+E S +K + KR P TP +  
 Sbjct: 534 KKEDKETSARKGKEITGVKRQSKSEHETSDAKKSVEDRGKRCP--TPEI 580

Score = 42 (6.3 bits), Expect = 5.7e-22, Sum P(2) = 5.7e-22  
 Identities = 13/58 (22%), Positives = 27/58 (46%)

Query: 167 EQSPIGRKKQEFPQEKIVVKREVKKRRQELEY-QVKAEELESMPPRKSLRLQGLAQST 223  
 EQ +KKQE E +++ K+ ++ E Q K E + ++ + G+ ++  
 Sbjct: 498 EQEQKAKKKQEAENEITEKQKQAKEEMKKESEQAKKEDKETSARKGKEITGVKRQS 555

Pedant information for DKFZphtes3\_27d1, frame 2

## Report for DKFZphtes3\_27d1.2

[LENGTH] 712  
 [MW] 81155.71  
 [pI] 8.21  
 [HOMOL] SWISSPROT:UBPB\_HUMAN UBIQUITIN CARBOXYL-TERMINAL HYDROLASE 11 (EC 3.1.2.15)  
 (UBIQUITIN THIOLESTERASE 11) (UBIQUITIN-SPECIFIC PROCESSING PROTEASE 13) (DEUBIQUITINATING  
 ENZYME 11) (KIAA0055). 4e-32  
 [FUNCAT] 06.13.01 cytoplasmic degradation [S. cerevisiae, YMR223w] 5e-33  
 [FUNCAT] 06.07 protein modification (glycosylation, acylation, myristylation,  
 palmitoylation, farnesylation and processing) [S. cerevisiae, YMR223w] 5e-33

[FUNCAT] 06.13 proteolysis [S. cerevisiae, YBL067c] 3e-19  
 [FUNCAT] 10.03.99 other osmosensing activities [S. cerevisiae, YDR069c] 2e-17  
 [FUNCAT] 03.10 sporulation and germination [S. cerevisiae, YDR069c] 2e-17  
 [FUNCAT] 30.10 nuclear organization [S. cerevisiae, YDR069c] 2e-17  
 [FUNCAT] 30.03 organization of cytoplasm [S. cerevisiae, YDR069c] 2e-17  
 [FUNCAT] 09.25 vacuolar and lysosomal biogenesis [S. cerevisiae, YDR069c] 2e-17  
 [FUNCAT] 04.05.01.04 transcriptional control [S. cerevisiae, YNL186w] 4e-17  
 [FUNCAT] 99 unclassified proteins [S. cerevisiae, YHL010c] 3e-12  
 [BLOCKS] BL00970A Nuclear transition protein 2 proteins  
 [BLOCKS] BL00972D  
 [BLOCKS] BL00972C  
 [BLOCKS] BL00972B  
 [BLOCKS] BL00972A  
 [EC] 3.1.2.15 Ubiquitin thiolesterase 5e-06  
 [PIRKW] alternative splicing 2e-11  
 [PIRKW] thiolester hydrolase 5e-06  
 [PIRKW] hydrolase 1e-14  
 [SUPFAM] RING finger homology 7e-11  
 [SUPFAM] deubiquinating enzyme SSV7 5e-16  
 [PROSITE] MYRISTYL 5  
 [PROSITE] AMIDATION 2  
 [PROSITE] CAMP\_PHOSPHO\_SITE 1  
 [PROSITE] CK2\_PHOSPHO\_SITE 10  
 [PROSITE] TYR\_PHOSPHO\_SITE 2  
 [PROSITE] UCH\_2\_2\_1  
 [PROSITE] PKC\_PHOSPHO\_SITE 17  
 [PROSITE] ASN\_GLYCOSYLATION 4  
 [PROSITE] UCH\_2\_1\_1  
 [PFAM] Ubiquitin carboxyl-terminal hydrolases family 2  
 [PFAM] Ubiquitin carboxyl-terminal hydrolases family 2  
 [KW] Alpha Beta  
 [KW] LOW\_COMPLEXITY 4.92 %

SEQ MLAMDTCKHVQLQLAQDHSSLPQKWHCVDCNTTESIWACLSCSHVACGRYIEEHALKH  
 SEG .....  
 PRD cccccccccchhhhhhhccccccccccccccccccccccccccccccccccccchhhhhhhhhhh

SEQ FQESSHPVALEVNMVFCYLCDDYVLNDNATGDLKLLRRLTSAIKSQNYHCTTRSGRFL  
 SEG .....  
 PRD hhhhhccccccccccccccccccccccccccccccccccccchhhhhhhhhhhhhcccccccccccc

SEQ RSMGTGDDSYFLHDGAQSLLQSEDQLYTALWHRRRILMGKIFRTWFEQSPIGRKKQEFPF  
 SEG .....  
 PRD cccccccccccccccccchhh

SEQ QEKIVVKREVKKRQLELEYQVKAELSMPPRKSRLQLGQAQSTIIIEIVSVQVPAQTPASP  
 SEG xxxxxxxxxxxxxxxxxxxx.....  
 PRD hheehhh

SEQ AKDKVLSTSENEISQKVSDDSVKRRPIVTPGVTLRLNLGTCYMNVLQVLSHLIFRQC  
 SEG .....  
 PRD cchhhhhhhhhhhhhhhhh

SEQ FLKLDLNLQWAMTASEKTRSCKHPPVTDTVVYQMNECQEKDTGFVCSRQSSLSGLSGGA  
 SEG .....xxxxxxxxxxxxxxxxx.....  
 PRD hhh

SEQ SKGRKMLIQPKEPTSYISLCHLHTLFQVMWSGKWALVSPFAMLSVWRILIPAFRGYA  
 SEG xxxxx.....  
 PRD cccccccccccccchhh

SEQ QQDAQEFLCELLDKIQRELETTGTSLPALIPTSRKLIKQVLNVVNNIFHGQLLSQVTCL  
 SEG .....  
 PRD hhh

SEQ ACDNKSNTIEPFWDLSLEFPERYQCSGKDIAQPCLVTEMLAKFTETEALEGKIYVCDQC  
 SEG .....  
 PRD cchhhhhhhhhhhhhhhhhhh

SEQ NSKRRRFSSKPVVLTEAQQLMICHLPQVLRHLKRFWRSGRNNREKIGVHVGFEEILNM  
 SEG .....  
 PRD cccccccccchhh

SEQ EPYCCRETLSLRPECFIYDLSAVVMHKGFGSGHYTAYCYNSEGGFWVHCNDSKLSMC  
 SEG .....  
 PRD ccc

SEQ TMDEVCKAQAYILFYTRVTENGHSLKLLPELLLSQHPNEDADTSSNEILS  
 SEG .....  
 PRD cchhh

## Prosite for DKFZphtes3\_27d1.2

PS000001	33->37	ASN_GLYCOSYLATION	PDOC00001
PS000001	90->94	ASN_GLYCOSYLATION	PDOC00001
PS000001	484->488	ASN_GLYCOSYLATION	PDOC00001
PS000001	653->657	ASN_GLYCOSYLATION	PDOC00001
PS000004	545->549	CAMP_PHOSPHO_SITE	PDOC00004
PS000005	6->9	PKC_PHOSPHO_SITE	PDOC00005
PS000005	113->116	PKC_PHOSPHO_SITE	PDOC00005
PS000005	116->119	PKC_PHOSPHO_SITE	PDOC00005
PS000005	213->216	PKC_PHOSPHO_SITE	PDOC00005
PS000005	254->257	PKC_PHOSPHO_SITE	PDOC00005
PS000005	261->264	PKC_PHOSPHO_SITE	PDOC00005
PS000005	315->318	PKC_PHOSPHO_SITE	PDOC00005
PS000005	320->323	PKC_PHOSPHO_SITE	PDOC00005
PS000005	394->397	PKC_PHOSPHO_SITE	PDOC00005
PS000005	453->456	PKC_PHOSPHO_SITE	PDOC00005
PS000005	506->509	PKC_PHOSPHO_SITE	PDOC00005
PS000005	542->545	PKC_PHOSPHO_SITE	PDOC00005
PS000005	548->551	PKC_PHOSPHO_SITE	PDOC00005
PS000005	580->583	PKC_PHOSPHO_SITE	PDOC00005
PS000005	608->611	PKC_PHOSPHO_SITE	PDOC00005
PS000005	611->614	PKC_PHOSPHO_SITE	PDOC00005
PS000005	676->679	PKC_PHOSPHO_SITE	PDOC00005
PS000006	125->129	CK2_PHOSPHO_SITE	PDOC00006
PS000006	164->168	CK2_PHOSPHO_SITE	PDOC00006
PS000006	223->227	CK2_PHOSPHO_SITE	PDOC00006
PS000006	247->251	CK2_PHOSPHO_SITE	PDOC00006
PS000006	249->253	CK2_PHOSPHO_SITE	PDOC00006
PS000006	313->317	CK2_PHOSPHO_SITE	PDOC00006
PS000006	506->510	CK2_PHOSPHO_SITE	PDOC00006
PS000006	525->529	CK2_PHOSPHO_SITE	PDOC00006
PS000006	661->665	CK2_PHOSPHO_SITE	PDOC00006
PS000006	706->710	CK2_PHOSPHO_SITE	PDOC00006
PS000007	193->200	TYR_PHOSPHO_SITE	PDOC00007
PS000007	192->200	TYR_PHOSPHO_SITE	PDOC00007
PS000008	218->224	MYRISTYL	PDOC00008
PS000008	355->361	MYRISTYL	PDOC00008
PS000008	359->365	MYRISTYL	PDOC00008
PS000008	471->477	MYRISTYL	PDOC00008
PS000008	589->595	MYRISTYL	PDOC00008
PS000009	171->175	AMIDATION	PDOC00009
PS000009	362->366	AMIDATION	PDOC00009
PS00972	274->290	UCH_2_1	PDOC00750
PS00973	619->638	UCH_2_2	PDOC00750

## Pfam for DKFZphtes3\_27d1.2

HMM_NAME	Ubiquitin carboxyl-terminal hydrolases family 2		
HMM	*GIqNlGNTCYMNSIIQCL*		
	G++NLGNTCYMNS++Q+L		
Query	274	GLRNLGNTCYMNSVLQVL	291
HMM_NAME	Ubiquitin carboxyl-terminal hydrolases family 2		
HMM	*YdLYgVICHYGntldyGHYWayVKNenhHRWkWYYFDDEtV*		
	YDL +V+ H+G + ++GHY+AY++N + ++W+ +D++		
Query	619	YDLSAVVMHHGKGFGSGHYTAYCYNSE--GGFWVHCNDSKL	657

DKFZphtes3\_27k4

group: transmembrane protein

Summary DKFZphtes3\_27k4 encodes a novel 490 amino acid protein with similarity to two hypothetical *C.elegans* proteins.

The novel protein contains 10 transmembrane regions and a leucine zipper. It is a member of the new 10 trans-membrane domain containing protein family which is specific for multicellular eukariotes.

No informative BLAST results; No predictive prosite, pfam or SCOP motife.

The new protein can find application in studying the expression profile of testis-specific genes and as a new marker for testicular cells.

strong similarity to *C.elegans* K07H8.2/ZK185.2  
membrane regions: 10

complete cDNA, complete cds potential start at Bp 109, few EST hits

Sequenced by GBF

Locus: unknown

Insert length: 1901 bp

Poly A stretch at pos. 1866, no polyadenylation signal found

```
1 GTGATTTACC AGAAAAACCA AGAAGACAGG CACAAAAAG CAAACGGCAT
51 TTGGCAAGAT GGATTATCAA CTGCAGTACA GACTTTTAGT AATAGATCTG
101 AGCAACACAT GGAGTATCAC AGTTTCTCAG AGCAGTCTTT TCATGCCAAT
151 AATGGGCACG CATCATCAAG CTGCAGCCAA AAGTATGATG ACTATGCCAA
201 TTATAATTAC TGTGATGGAA GGGAGACTTC AGAAACCACT GCCATGTTAC
251 AAGATGAAGA TATATCTAGT GATGGTGATG AAGATGCTAT TGTAGAAGTG
301 ACCCCAAAAT TACCAAAGGA ATCCAGTGGC ATCATGGCAT TGCAAATACT
351 TGTGCCCTTT TTGCTAGCTG GTTTTGGAA CAGTTTCAGCT GGCATGGTAC
401 TGGATATAGT ACAGCACTGG GAGGTGTTCA GAAAAGTTAC AGAAGTTTC
451 ATTTTAGTCC CTGCACTTCT TGGTCTCAA GGGAACTGG AAATGACATT
501 GGCATCCAGA TTATCCACTG CAGTAAATAT TGGGAAGATG GATTCAACCA
551 TTGAAAAGTG GAACCTAATA ATTGGCAACT TGGCTTTAAA GCAGGTTTCA
601 GCAACAGTAG TGGGTTTCTC AGCAGCTGTG GCAGCAATTA TATTGGGCTG
651 GATTCCAGAA GGAAATATT ACCTTGATCA TTCCATACTT CTGTGCTCTA
701 GCAGTGTGGC AACTGCCTTC ATTGCATCTC TTCTGCAGGG AATAATAATG
751 GTTGGGGTTA TCGTTGGTTC AAAGAAGACT GGTATAAATC CTGATAATGT
801 TGCTACACCC ATTGCTGCTA GTTTTGGCGA CCTTATAACT CTTGCCATAT
851 TGGCTTGGAT AAGTCAGGGC TTATCTCTCT GTCTTGAGAC CTATTACTAC
901 ATTTCTCCAT TAGTTGGTGT ATTTTCTTGT GCTCTAACCC CTATTGGTAT
951 TATAATAGCT GCCAAACATC CAGCCACAAG AACAGTTCTC CACTCAGGCT
1001 GGGAGCCTGT CATAACAGCT ATGCTTATAA GTAGCATTTG GGGCCTTATT
1051 CTGGACACAA CTGTATCAGA CCCAAACTTG GTTGGGATTG TTGTTTACAC
1101 GCCAGTTATT AATGGTATTG TTGGTAATTT GGTGGCCATT CAGGCTAGCA
1151 GGATTCTTAC CTACCTCCAT TTACATAGCA TTCCAGGAGA ATTGCCTGAT
1201 GAACCCAAAG GTTGTACTA CCCATTAGA ACTTCTTTG GTCCAGGAGT
1251 AAATAATAAG TCTGCTCAAG TTCTACTGCT TTTAGTGATT CCTGGACATT
1301 TAATTTTCTT CTACACTATT CATTTGATGA AAAGTGGTCA TACTTCTTTA
1351 ACTATAATCT TCATAGTAGT GTATTTATTT GGCCTGTGTT TACAGGTATT
1401 TACCTTGCTG TGGATTGCTG ACTGGATGGT CCATCACTTC TGGAGGAAAG
1451 GAAAGGACCC GGATAGTTTC TCCATCCCTT ACCTAACAGC ATTGGGTGAT
1501 CTGCTCGGGA CAGCTCTGTT AGCCTTAAGT TTTCAATTTT TTTGGCTTAT
1551 TGGAGATCGA GATGGAGATG TTGGAGACTA ATAAATCTTA CAACTGCTTC
1601 TCAAGTTACC AAGGAAGAAA ATACACGACA ACCACTTATG GCTCTTTTTC
1651 AAAACTCTTA AATCAGTAGT TTGACTTTTG CCAGGGTAAT CTTCAAGTTGG
1701 CCCTGATTCA ATTAATGGC CTTAATTTTT TTTTAAGGAA TTTGTGTCAA
1751 AACCAGAAATG AAGAGTATTC GTGCTGCTTT TCATAGAATA AATGATAATT
1801 TGACATAGAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA
1851 AAAAAAAAAA AAGGGGAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA
1901 G
```

## BLAST Results

No BLAST result

## Medline entries

No Medline entry

Peptide information for frame 1

ORF from 109 bp to 1578 bp; peptide length: 490  
Category: similarity to unknown protein

```

1 MEYHSFSEQS FHANNGHASS SCSQKYDDYA NYNYCDGRET SETTAMLQDE
51 DISSDGDDEDA IVEVTPKLPK ESSGIMALQI LVPFLLAGFG TVSAGMVLDI
101 VQHWVEVFRKV TEVFILVPAL LGLKGNLEMT LASRLSTAVN IGMDSPIEK
151 WNLIIGNLAL KQVQATVVGF LAAVAAILG WIPEGKYLD HSILLCSSSV
201 ATAFIASLLQ GIIMVGIVG SKKTGINPDN VATPIAASFG DLITLAILAW
251 ISQGLYSCLC TYYIISPLVG VFFLALTPIW IIIAAKHPAT RTVLHSGWEP
301 VITAMVISSI GGLILOTTVS DPNLVGIVVY TPVINGIGGN LVAIQASRIS
351 TYLHLHSIPG ELPDEPKGCY YPFRTFFGPG VNNKSAQVLL LLVIPGHLIF
401 LYTIIHLMKSG HTSLTIIFIV VYLFQAVLQV FTLLWIADWM VHHFWRKGKD
451 PDSFSIPYLT ALGDLGTAL LALSFLHLWL IGDROGDVGD

```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphtes3\_27k4, frame 1

TREMBL:AF036704 2 gene: "ZK185.2"; Caenorhabditis elegans cosmid  
ZK185., N = 1, Score = 730, P = 3.1e-72

TREMBL:AF047659 9 gene: "K07H8.2"; Caenorhabditis elegans cosmid  
K07H8., N = 1, Score = 940, P = 1.7e-94

>TREMBL:AF047659 9 gene: "K07H8.2"; Caenorhabditis elegans cosmid K07H8.  
Length = 507

HSPs:

Score = 940 (141.0 bits), Expect = 1.7e-94, P = 1.7e-94  
Identities = 204/412 (49%), Positives = 271/412 (65%)

```

Query: 68 LPKESSGIMALQILVPFLLAGFGTVSAGMVDIVQHWVEVFRKVEVFILVPALLGLKGNL 127
      +P ESS ++ Q+L PF +AG G V AG+VL IV W +F ++ E+ ILVPALLGLKGNL
Sbjct: 82 IPAESYVLFQVLFPPFAVAGLGMVFAGLVLSIVVTWPLFEEIPEILILVPALLGLKGNL 141

Query: 128 EMTLASRLSTAVNIGKMDSPIEKNWLIIGNLALKQVQATVVGFLLAAVAAILGWIPEGKY 187
      EMTLASRLST N+G MDS ++ ++I NLAL QVQATVV FLA+ A L +IP G +
Sbjct: 142 EMTLASRLSTLANLGHMDSSKQRKDVVIANLALVQVQATVVAFASFAAALAFIPSGDF 201

Query: 188 YLDHSILLCSSSVATAFIASLLQGIIMVGIVGSKKTGINPDNVATPIAASFGDLITLAI 247
      H L+C+SS+ATA ASL+ ++MV VIV S+K INPDNVATPIAAS GDL TL +
Sbjct: 202 DWAHGALMCASSLATACSASLVLSLLMVVIVTSRKYNINPDNVATPIAASLGDLTTLTV 261

Query: 248 LAWISQGLYSCLCTYYIISPLVG VFFLALTPIWIIIAAKHPATRTVLHSGWEPVITAMVI 307
      LA+ T +++ +V V FL L P WI IA ++ T+ L+GW PVI +M+I
Sbjct: 262 LAFFGSVFLKAHNTESWLNIVIVLFLLLLPFFWIKIANENEGTQETLYNGWTPVIMSMLI 321

Query: 308 SSIGGLILOTTVSDPNLVGIVVYTPVINGIGGNLVAIQASRISTYLHLHSIPGELPDEPK 367
      SS GG IL+T V + + Y PV+NG+GGNL A+QASR+STY H G LP+E
Sbjct: 322 SSAGGFILETAVRRYH--SLSTYGFVLNGVGGNLAQVQASRLSTYFHKAGTVGVLPNEWT 379

Query: 368 GCYYPF--RTFFGPGVNNKSAQVLLLVIPGHLIFLYTIHLM---KSGHTSLTIIFIVV 421
      + R FF ++SA+VLLLLV+PGH+ F + I L K+ T +F +
Sbjct: 380 VSRFTSVQRAFFSKEWDSRSARVLLLVVPGHICFNFLIQLFTLSKNNVTPHGPLEFTSL 439

Query: 422 YLFQAVLQVFTLLWIADWMVHHFWRKGKDPDSFSIPYLTALGDLGTALLALSF 475
      Y+ A++QV LL++ +V W+ DPD+ IPYLTALGDLGT LL + F
Sbjct: 440 YMIAAIIQVVILLFVCQLLVALLWKWKIDPDNSVIPYLTALGDLGTGLLFIVF 493

```

Pendant information for DKFZphtes3\_27k4, frame 1

Report for DKFZphtes3\_27k4.1

[LENGTH] 490  
[MW] 53266.39



```

SEQ      MEYHSFSEQSFHANNGHASSSCSQKYDDYANYNYCDGRETSETTAMLQDEIDSSDGEDA
SEG
PRD      cccccccceeeccccccccccccccccccccceeecccccccchhhhhhhhcccccccccee
MEM      .....

SEQ      IVEVTPKLPKESSGIMALQILVPFLLAGFTVSAGMVLDIVQHWEVFRKVTEVFI LVPAL
SEG
PRD      eeeeeccccccchhhhhhhhhhhhhhhhhccchhhhhhhhhcchhhhhccccceeeecccc
MEM      .....MMMMMMMMMMMMMMMMMMMMMMMMMMMMMM.....MMMMMMMMMMMMMMMM

SEQ      LGLKGNLEMTLASRLSTAVNIGKMDSPIEKWNLIIGNLALKQVQATVVGVFLAAVAAILG
SEG
PRD      ccccchhhhhhhhhhhhhhhcccccceeeehhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh
MEM      MMMMMM.....MMMMMMMMMMMMMMMMMMMM.....

SEQ      WIPEGKYLDHSILLCSSSVATAFIASLLQGIIMVGIVGSKKTGINPDNVATPIAASF
SEG
PRD      hccccceeeccccceehhhhhhhhhhhhhhhhhhhheeecccccccccccccccccccccc
MEM      .....MMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMM.....MMMMMMMM

SEQ      DLITLAILAWISQGLYSCLETYIYYISPVLGVFFLALTPIWIIIAAKHPATRTVLHSGWEP
SEG
PRD      cchhhhhhhhhhhhhhhccceeeehhhhhhhhhhhcchhhhhhhhhccccccccchhhhhhh
MEM      MMMMMMMMMMMMMMMM.....MMMMMMMMMMMMMMMMMMMM.....MMMMMM

SEQ      VITAMVISSIGGLILDTTVSDPNLVGIVVYTPVINGIGGNLVAIQASRISTYLHLHSIPG
SEG
PRD      hcchhhhhccceeeccccccccceeeecceeeccccceeeehhhhhhhhhhhccccc
MEM      MMMMMMMMMMMMMMMM.....

SEQ      ELPDEPKGCYYPFRTFFPGPVNNKSQAQVLLLLVIPGHLIFLYTIHLMKSGHTSLTIIFIV
SEG
PRD      cccccccccceeeccccchhhhhhhhhhhcccccchhhhhhhhhcccccceeeehhh
MEM      .....MMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMM.....MMMMMMMM

SEQ      VYLFGAVLQVF TLLWIADWMVHHFRWKGDPSFSIPYLTALGDLLGTALLALS FHF LWL
SEG
PRD      hhhhhhhhhhhhhhhhhhhhhhhhhhhhhcccccceeeecchhhhhhhhhhhheeee
MEM      MMMMMMMMMMMMMMMMMMMM.....MMMMMMMMMMMMMMMMMMMMMMMMMMMM

SEQ      IGDRDGVDVG
SEG
PRD      ecccccccc
MEM      MM

```

PS000001	383->387	ASN_GLYCOSYLATION	PDOC000001
PS000004	108->112	CAMP_PHOSPHO_SITE	PDOC000004
PS000005	23->26	PKC_PHOSPHO_SITE	PDOC000005
PS000005	65->68	PKC_PHOSPHO_SITE	PDOC000005
PS000005	221->224	PKC_PHOSPHO_SITE	PDOC000005
PS000006	5->9	CK2_PHOSPHO_SITE	PDOC000006
PS000006	54->58	CK2_PHOSPHO_SITE	PDOC000006
PS000006	146->150	CK2_PHOSPHO_SITE	PDOC000006
PS000006	238->242	CK2_PHOSPHO_SITE	PDOC000006
PS000006	257->261	CK2_PHOSPHO_SITE	PDOC000006
PS000006	296->300	CK2_PHOSPHO_SITE	PDOC000006
PS000006	318->322	CK2_PHOSPHO_SITE	PDOC000006
PS000007	25->33	TYR_PHOSPHO_SITE	PDOC000007
PS000008	90->96	MYRISTYL	PDOC000008
PS000008	122->128	MYRISTYL	PDOC000008
PS000008	216->222	MYRISTYL	PDOC000008
PS000008	220->226	MYRISTYL	PDOC000008

WO 01/12659

PCT/IB00/01496

PS00008	254->260	MYRISTYL	PDOC00008
PS00008	336->342	MYRISTYL	PDOC00008
PS00008	339->345	MYRISTYL	PDOC00008
PS00013	12->23	PROKAR_LIPOPROTEIN	PDOC00013
PS00013	248->259	PROKAR_LIPOPROTEIN	PDOC00013
PS00029	459->481	LEUCINE_ZIPPER	PDOC00029

(No Pfam data available for DKFZphtes3\_27k4.1)

DKFZphtes3\_27014

group: testes derived

DKFZphtes3\_27014 encodes a novel 358 amino acid protein with similarity to C. elegans cosmid C55A6.

The new protein contains a C3HC4 zinc finger (RING finger) signature. The ring finger structure binds two atoms of zinc, and is involved in mediating protein-protein interactions. No informative BLAST results; No predictive prosite, pfam or SCOP motif.

The new protein can find application in studying the expression profile of testis-specific genes.

similarity to C.elegans C55A6.1

complete cDNA, complete cds, EST hits

Sequenced by GBF

Locus: /map="6"

Insert length: 2158 bp

Poly A stretch at pos. 2137, polyadenylation signal at pos. 2120

```
1 CCGAGGCCAG AGAGAAAAGA CTGCGAGGTG GCCGCAGCTG TGGCCGGAGA
51 GCACAAAGAA TGAACCAAGCA GTGGAAGAGA AAATACTGTA AGCTGGCTGA
101 CTGCTGGTGA AGAAAATGCT TTATTTTGTG GGCAGGCATC TGTGGGATCT
151 GTAATAGAAA TATATTGGAG TAATTCAPGA TTCTGTGGTT GGCCCTTTTG
201 ACTGCTCTCT CTACAGGTTT AATTGGGCA TTTACTCATT TTCTATGGCTC
251 CAAGGACCAT GTATGTGTTG GGGATCTTCA ATATTCATGT TATTTCTCC
301 TTTGGTCTTA TATGATTGTT ACCTTTATGA AGCTTTAGTG ATTACAAAGC
351 ACTTTTCTTG TCCATTTTTC CCTGAGCTTT GTAAACTCTG ATTTGCAGGA
401 TGGCTGGCTG TGGTGAAATT GATCATTCAA TAAACATGCT TCCTACAAAC
451 AGGAAAGCGA ACGAGTCTCT TTCTAATACT GCACCTTCTT TAACCGTCCC
501 TGAATGTGCC ATTTGTCTGC AAACATGTGT TCATCCAGTC AGTCTGCCCT
551 GTAAGCACGT TTTCTGCTAT CTATGTGTAA AAGGAGCTTC ATGGCTTGGA
601 AAGCGGTGTG CTCTTTGTCT ACAAGAAATT CCGGAGGATT TCCTTGACAA
651 GCCAACCTTG TTGTCAACAG AAGAAGCTCA GGCAGCAAGT AGAGGAAATG
701 GTGAATATGC ATGGTATTAT GAAGGAAGAA ATGGGTGGTG GCAGTACGAT
751 GAGCGCACTA GTAGAGAGCT GGAAGATGCT TTTTCCAAAG GTAAAAAGAA
801 CACTGAAATG TTAATTGCTG GCTTTCTGTA TGTGCTGAT CTGAAAACA
851 TGGTTCAATA TAGGAGAAAT GAACATGGAC GTCCGAGGAA GATTAAAGCA
901 GATATAATAG ATATACCAAA GAAGGGAGTA GCTGGACTTA GGCTAGACTG
951 TGATGCTAAT ACCGTAAACC TAGCAAGAGA GAGCTCTGCT GACGGAGCGG
1001 ACAGTGTATC AGCACAGAGT GGAGCTTCTG TTCAGCCCTT AGTGTCTTCT
1051 GTAAGGCCCC TAACATCAGT AGATGGTCAG TTAACAAGCC CTGCAACACC
1101 ATCCCCGTGAT GCAAGCACTT CTCTGGAAGA CTCTTTTGCT CATTTACAC
1151 TCAGTGGAGA CAACACAGCT GAAAGGAGTC ATAGGGGAGA AGGAGAAGAA
1201 GATCATGAAT CACCATCTTC AGGCAGGGTA CCAGCACCAG ACACCTCCAT
1251 TGAAGAAACT GAATCAGATG CCAGTAGTGA TAGTGAGGAT GTATCTGCAG
1301 TTGTTGACCA GCACTCCTTG ACCCAACAGA GACTTTTGGT TTCTAATGCA
1351 AACCAAGACG TACCCGATCG ATCAGATCGA TCGGGAAGTG ATCGATCAGT
1401 AGCAGGGGGT GGAACAGTGA GTGTCACTGT CAGATCTAGA AGCCCTGATG
1451 GACAGTGCAC AGTAACTGAA GTTTAAATAA AAATGTCTTC AGCTCCATGC
1501 TCAAGGTTGA AAGGGTTACC TGTAATTTTC TGCCACATA ACATTATACT
1551 CATCCCTAGT AGTGCAATTT GGGAGTTGGG GTGGGAAGGG GTATGGGAAG
1601 GATAGACTCA TAATTTAAAT GTCTAACATG TCTCTGTTGA GAAATTTATT
1651 TAATGTAAGG AACTTGGGTG TTAATAGTTG AGAGCTGTTT AGTAATAACC
1701 CAGTTTCTT GAGGTCTGTT TACTTTATAC TTTTAAAAA CTCTGTAGT
1751 TCTTTTGCCC AGTGTGTTTG TATTATCTGT GCATTAAATG TCCTCATCTG
1801 ACTCCTGCAT TGTGCTTAT TTTCTGCAT GGATTGGCAT AAGACCATTA
1851 CTAAATTTG GCACCTGTGA GATGTTTGAT ATTATGAACA GGAAACATAA
1901 TTTAATGTAT GAATAGATGT GAATTTGGGA TTTCAAAATA GATGAATAAC
1951 AACATATTTA TAGTAAAGTT ATTGAAATGG AAATGAAAAC AGCCAGTAAC
2001 TTATGTTTCA GAATGTTTGT AACACACTTC ATGGTGTTC CATAGCCTTT
2051 GCTGTCTAGT CTTATAGTTT GAGGTTTTTT TGGTCTGCAT TTTTCTTTT
2101 GATTACAAAA TTTATAATTT AATAAATACT AGAGTTTATC AAAAAAAG
2151 AAAAAAAG
```

## BLAST Results

Entry HSG117 from database EMBL:

human STS SHGC-36270.

Score = 1148, P = 8.9e-45, identities = 240/250

## Medline entries

No Medline entry

## Peptide information for frame 1

ORF from 400 bp to 1473 bp: peptide length: 358  
Category: similarity to unknown protein  
Prosites motifs: ZINC\_FINGER\_C3HC4 (51-61)

```
1 MAGCGEIDHS INMLPTNRKA NESCSNTAPS LTVPECAICL QTCVHPVSLP
51 CKHVFCYLCV KGASWLGKRC ALCRQEIPED FLDKPTLLSP EELKAASRGN
101 GEYAWYYEGR NGWWQYDERT SRELEDAFSK GKNTEMLIA GFLYVADLEN
151 MVQYRRNEHG RRRKIKRDII DIPKGVAGL RLDCDANTVN LARESSADGA
201 DSVSAQSGAS VQPLVSSVRP LTSVDGQLTS PATSPDAST SLEDSFAHLQ
251 LSGDNTAERS HRGEGEDHE SPSSGRVPAP DTSIEETESD ASSDSEDVSA
301 VVAQHSILTQ RLLVSNANQT VPDRSDRSRG DRSVAGGGTV SVSVRSRRPD
351 GQCTVTEV
```

## BLASTP hits

No BLASTP hits available

## Alert BLASTP hits for DKFZphtes3\_27o14, frame 1

TREMBL:CEC55A6\_1 gene: "C55A6.1"; Caenorhabditis elegans cosmid C55A6,  
N = 2, Score = 165, P = 4.2e-15

SWISSPROT:YW26\_CAEEL HYPOTHETICAL 39.3 KD PROTEIN C02B8.6 IN CHROMOSOME  
X., N = 2, Score = 136, P = 3.1e-11

>TREMBL:CEC55A6\_1 gene: "C55A6.1"; Caenorhabditis elegans cosmid C55A6  
Length = 484

## HSPs:

Score = 165 (24.8 bits), Expect = 4.2e-15, Sum P(2) = 4.2e-15  
Identities = 42/106 (39%), Positives = 61/106 (57%)

Query: 75 QEIPEDFLDKPTLLSPEELKAASRGNGEYAWYYEGRN-GWWQYDERTSRELEDAFSKGGK 133  
Q +P LD ++ PEE K Y W Y G+N GWW+++ R RE+E+A++ GK  
Sbjct: 93 QNVPALDLDA-SICDPEERK-----Y-WIYSGKNQGWWRFEPRNEREIEEAYNAGKC 142

Query: 134 NTEMLIAGFLYVADLENMVQYRRNEHGRRRRIKR---DIID-IPKKGAVAGL 180  
+ E++I G YV D +QY R + R +KR D D I KG+AG+  
Sbjct: 143 HCEVVICGRPYVIDFHQFLQYPRGVPNQARHVKRVSADDFDGIGVKGLAGI 193

Score = 96 (14.4 bits), Expect = 4.2e-15, Sum P(2) = 4.2e-15  
Identities = 19/54 (35%), Positives = 30/54 (55%)

Query: 35 ECAICLQTCVHPVSLP-CKHVFCYLCVKGASW--LGKRCALCRQEIPEDFLDKPT 86  
EC IC + P ++P C H FC++C+KG +G C +CR I + +P+  
Sbjct: 11 ECPICQCKMIVPTTIPACGHKFCFICLKGVMNDMGG-CPMCRGPIDSNIFAQPS 64

## Pedant information for DKFZphtes3\_27o14, frame 1

## Report for DKFZphtes3\_27o14.1

[LENGTH] 358  
[MW] 38818.90  
[pI] 5.17  
[HOMOL] TREMBL:CEC55A6\_1 gene: "C55A6.1"; Caenorhabditis elegans cosmid C55A6 2e-12  
[FUNCAT] 11.04 dna repair (direct repair, base excision repair and nucleotide excision  
repair) [S. cerevisiae, YCR066w] 3e-04  
[FUNCAT] 03.19 recombination and dna repair [S. cerevisiae, YCR066w] 3e-04  
[FUNCAT] 30.10 nuclear organization [S. cerevisiae, YCR066w] 3e-04

```

[FUNCAT] 06.07 protein modification (glycosylation, acylation, myristylation,
palmitoylation, farnesylation and processing) [S. cerevisiae, YCR066w] 3e-04
[FUNCAT] 06.10 assembly of protein complexes [S. cerevisiae, YDR265w] 4e-04
[FUNCAT] 30.19 peroxisomal organization [S. cerevisiae, YDR265w] 4e-04
[BLOCKS] BL00518 Zinc finger, C3HC4 type, proteins
[PROSITE] MYRISTYL 2
[PROSITE] AMIDATION 3
[PROSITE] CAMP_PHOSPHO_SITE 1
[PROSITE] CK2_PHOSPHO_SITE 12
[PROSITE] TYR_PHOSPHO_SITE 1
[PROSITE] ZINC_FINGER_C3HC4 1
[PROSITE] PKC_PHOSPHO_SITE 9
[PROSITE] ASN_GLYCOSYLATION 2
[PFAM] Zinc finger, C3HC4 type (RING finger)
[KW] Irregular
[KW] 3D
[KW] LOW_COMPLEXITY 19.83 %

```

```

SEQ MAGCGEIDHSINMLPTNRKANESCSNTAPSLTVPECAICLQTCVHPVSLPCKHVFCYLCV
SEG .....
lrmd- .....TTTTTEETTTEETTTTEEEHHHHH

SEQ KGASWLGKRCALCRQEIPEDFLDKPTLLSPEELKAASRGNGEYAWYYEGRNGWQYDERT
SEG .....
lrmd- HHHHHHCCBTTTTCBCGGG-CBCC.....

SEQ SRELEDAFSKGGKNTMLIAGFLYVADLENMVQYRRNEHGRRRRIKRDIIIPKKGVAGL
SEG .....XXXXXXXXXXXXXXXXX.....
lrmd- .....

SEQ RLDCDANTVNLARESSADGADSVSAQSGASVQPLVSSVRPLTSVDGQLTSPATPSPDAST
SEG .....XXXXXXXXXXXXXXXXX.....
lrmd- .....

SEQ SLEDSFAHLQLSGDNTAERSHRGEGEEDHESPSSGRVPAPDTSIEETSDASSDSEDVSA
SEG X.....XXXXXXXXXXXXXXXXXXXXX.....
lrmd- .....

SEQ VVAQHSILTQQRLLVSNANQTVPDSDRSRGTDSDSVAGGTVSVSVRRRPDGGQCTVTEV
SEG xxx.....XXXXXXXXXXXXXXXXXXXXX.....
lrmd- .....

```

## Prosite for DKFZphtes3\_27ol4.1

PS00001	21->25	ASN_GLYCOSYLATION	PDOC00001
PS00001	318->322	ASN_GLYCOSYLATION	PDOC00001
PS00004	132->136	CAMP_PHOSPHO_SITE	PDOC00004
PS00005	16->19	PKC_PHOSPHO_SITE	PDOC00005
PS00005	120->123	PKC_PHOSPHO_SITE	PDOC00005
PS00005	217->220	PKC_PHOSPHO_SITE	PDOC00005
PS00005	260->263	PKC_PHOSPHO_SITE	PDOC00005
PS00005	274->277	PKC_PHOSPHO_SITE	PDOC00005
PS00005	325->328	PKC_PHOSPHO_SITE	PDOC00005
PS00005	330->333	PKC_PHOSPHO_SITE	PDOC00005
PS00005	343->346	PKC_PHOSPHO_SITE	PDOC00005
PS00005	346->349	PKC_PHOSPHO_SITE	PDOC00005
PS00006	32->36	CK2_PHOSPHO_SITE	PDOC00006
PS00006	89->93	CK2_PHOSPHO_SITE	PDOC00006
PS00006	120->124	CK2_PHOSPHO_SITE	PDOC00006
PS00006	195->199	CK2_PHOSPHO_SITE	PDOC00006
PS00006	222->226	CK2_PHOSPHO_SITE	PDOC00006
PS00006	240->244	CK2_PHOSPHO_SITE	PDOC00006
PS00006	282->286	CK2_PHOSPHO_SITE	PDOC00006
PS00006	287->291	CK2_PHOSPHO_SITE	PDOC00006
PS00006	293->297	CK2_PHOSPHO_SITE	PDOC00006
PS00006	320->324	CK2_PHOSPHO_SITE	PDOC00006
PS00006	328->332	CK2_PHOSPHO_SITE	PDOC00006
PS00006	354->358	CK2_PHOSPHO_SITE	PDOC00006
PS00007	98->107	TYR_PHOSPHO_SITE	PDOC00007
PS00008	329->335	MYRISTYL	PDOC00008
PS00008	337->343	MYRISTYL	PDOC00008
PS00009	66->70	AMIDATION	PDOC00009
PS00009	130->134	AMIDATION	PDOC00009
PS00009	159->163	AMIDATION	PDOC00009
PS00518	51->61	ZINC_FINGER_C3HC4	PDOC00449

Pfam for DKFZphtes3\_27o14.1

HMM_NAME	Zinc finger, C3HC4 type (RING finger)		
HMM	*CPICFCTFQlDyPWPfdePmMlPCGhSFCypCIrrW.....CPmC*		
	C+IC	L + P++LPC+H+FCY C++	C +C
Query	36	CAIC-----LQT---CVHPVSLPCKHVFCYLCVKGASWLKGKRCALC	73

DKF2phtes3\_28d14  
-----

group: testes derived

DKF2phtes3\_28d14 encodes a novel 97 amino acid protein without similarity to known proteins.

No informative BLAST results; No predictive prosite, pfam or SCOP motife.

The new protein can find application in studying the expression profile of testis-specific genes.

unknown

complete cDNA, complete cds, EST hits

Sequenced by GBF

Locus: unknown

Insert length: 1279 bp

Poly A stretch at pos. 1232, no polyadenylation signal found

```
1 GGAGCTCAGA AGTTGGGCAA AGGTCACAGC AGACTTCCTG AAAAGCAGAC
51 ACTGAGGAAC ACAGTGGAGA GCGGGAGTTC ACAGCGACGC AGCTGAGGAC
101 GACGCAGGAC CTCTCCCAAA GGTGCTGCAG CTCCAGCACC AGGGGCCAGG
151 GCTGCGCGCA CAGCAGCTCA GCAACCCTTG CTGTGCTCAA GTTCTTGGGG
201 ATTCAGAGCT AAGTTCAAAA TTAGAAACA GTGCCTTAAA GACGGGCAAG
251 AAAACCCGGT GTGGGAGTCT GCTCATCTAT GGTTTGTAC TGCTCTCGCT
301 TTGATATTCT TAAATTCCTA GGTACCAATG AAAAGCCAA GTGAACGTGG
351 CAGAGTGAGG AGGAGACAGG AGCGTGTGCA CCTCCATCT GTGAGAGGCA
401 CACTTCAGTC TGGGTTCAAG ATGCAGAATG GTGCCTACAG CAAAAAATAA
451 AAAAAACACC TCCTCCCTTC TTTACCATT GAATGGACAT TTTCTTACC
501 TGTGATCCCA ACAGAAACAG ATCCAGACCT ATCATGTGAA GTCCACGTTC
551 CAGGATCAGA AGTAACCAGT TTATGGACTG AGCTTACACG GGAAAGTCTA
601 CCCCCGACTC CTTCTGGATA GTAACATACA CAGCTGCATA AAAACGTCTC
651 CAAGGGGACA TACGATGCAT TTGCTTGGTG TCCAGCCAA GCTCCCCACC
701 GGCAGACTCA CTGTTCTTGA GAGCTCGAGA GCTCGTCTCC TATCAATCAG
751 AGAACCCCAT CAGCTGTGAC CAACAGAGCT GGAGCCCTCT GTGGAGGGAG
801 CTGACCCAC ACACAGGACA GAGCAGAATC CTGATTATTT TACAACTGCG
851 AAACCTTCTG AGTAAGAAGA CAAAAATATA CATTCCAAGG TATCTGTAAA
901 GTGCTTGGAA GATGCAGACA GCTGCACCGA GGGGCTCTGA TCCATCCACA
951 CGCTGCGGCT TGCTGCGGTC ACACACACGG TCTCAGTCAC GTGATGGTTT
1001 TGCTTTTATT TCTTAAACGG CTGAGTGATA ATCCAGCTAG TGTGCAGTCA
1051 TTTTCATACCT TTCAATGGGC GTCACCGCAG TGACGCTGCC CCAGCCCAT
1101 GCTGAGGGCC GACACAATTC ACGGAACAGA TTCATCATAT TTGGTCTTTA
1151 TGTAATAAT AAATGTTTAA AAATTGCCCTA AATATAAAAA AAAAAAAAAA
1201 AAAAAAAAAA AAAAAAAAAA AAAGGGCGGC CGAAAAAAAA AAAAAAAAAA
1251 AAAAAAAAAA AAAAAAAAAA GGGCGGCCG
```

BLAST Results  
-----

No BLAST result

Medline entries  
-----

No Medline entry

Peptide information for frame 1  
-----

ORF from 328 bp to 618 bp; peptide length: 97

Category: putative protein

```
1 MKKPSERGRV RRRQERVHLP SVRGTLQSGF KMONGAYSKK KNNTLLPSLP
51 FEWTFSLPVI PTETDPLSC EVHVPGEVST SLWTELTRES LPPTPSG
```

## BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphtes3\_28d14, frame 1

No Alert BLASTP hits found

Pedant information for DKFZphtes3\_28d14, frame 1

-----  
Report for DKFZphtes3\_28d14.1

{LENGTH} 97  
{MW} 10945.56  
{pI} 9.80  
{PROSITE} MYRISTYL 2  
{PROSITE} CAMP\_PHOSPHO\_SITE 2  
{PROSITE} CK2\_PHOSPHO\_SITE 2  
{PROSITE} PKC\_PHOSPHO\_SITE 3  
{KW} All\_Alpha  
{KW} LOW\_COMPLEXITY 12.37 %

SEQ MKKPSEGRVRRRQERVHLPSVRGTLQSGFKMQNGAYSKKKKNTLLPSLPFEWTFSLPVI  
SEG .....XX  
PRD cccccchhhhhhhhhhhcc

SEQ PTETDPDLSCEVHVPGEVTSWLTELTRSLPPTPSG  
SEG .....  
PRD cccccccceeeccccchhhhhhhhhhhcccccccc

Prosites for DKFZphtes3\_28d14.1

PS00004	2->6	CAMP_PHOSPHO_SITE	PDOC00004
PS00004	41->45	CAMP_PHOSPHO_SITE	PDOC00004
PS00005	5->8	PKC_PHOSPHO_SITE	PDOC00005
PS00005	21->24	PKC_PHOSPHO_SITE	PDOC00005
PS00005	38->41	PKC_PHOSPHO_SITE	PDOC00005
PS00006	62->66	CK2_PHOSPHO_SITE	PDOC00006
PS00006	64->68	CK2_PHOSPHO_SITE	PDOC00006
PS00008	24->30	MYRISTYL	PDOC00008
PS00008	76->82	MYRISTYL	PDOC00008

(No Pfam data available for DKFZphtes3\_28d14.1)



DKFZphtes3\_2a11

group: testes derived

DKFZphtes3\_2a11 encodes a novel 1048 amino acid protein with very weak similarity to mucins.

No informative BLAST results; No predictive prosite, pfam or SCOP motife.

The new protein can find application in studying the expression profile of testis-specific genes.

similarity to mucin

complete cDNA, complete cds, EST hits

Sequenced by EMBL

Locus: unknown

Insert length: 4082 bp

Poly A stretch at pos. 4060, polyadenylation signal at pos. 4034

```
1 GAGGACTGCG AGCACAGCGG CGGCCGGGTG GCGGGGGTGA GTGGGGCCAG
51 CGGGGCTGGA CAGCAGCGGG CCGCGGGGCG CGCGGGCGCG ATCCCTCCCC
101 GCGCGCGCGG AGCACATCGC CGCGCGCGAG ATGGGCGCTC CGCGGGCACCC
151 CCAGGCGCGG GAGATAGAAG CGGCGGGTGC GGGCGGGCGG CGCGGGCTAC
201 AGGTGGAAAT GAGTCTCAA CAGTTTCCTC GGTAGGAGC CCCTTCTACC
251 GGGCTGAGCC AGGCCCTTTC TCAGATTGCA AACAGTGGTT CTGCTGGATT
301 GATAAACCCA GCTGCTACAG TCAATGATGA ATCTGGTCCA GATTCTGAAG
351 TCAGTGCCAG GGAGCACATG AGTTCCAGCA GTCCTCTCCA GTCCCGGGAG
401 GAGAAGCAAG AGCCTGTGTG GGTAAAGGCC TATCCACAGG TGCAGATGTT
451 GTCGACACAC CATGCTGTGC CATCAGCCAC ACCTGTTGCA GTGACAGCCC
501 CGCGAGCACA CCTGACGCCA GCAGTGCCAC TTTCATTTTC GGAGGGACTT
551 ATGAAGCCCG CCGGAAGGCC CACCATGCCT AGCCGTCCCA TTGCTCCTGC
601 TCCACCTTCT ACCCTGTGAC TTCCCGCCAA GGTCCAGGG CAGGTTACCG
651 TTACCATGGA GAGTAGCATC CCTCAAGCTT CAGCCATTCC TGTGGCAACA
701 ATCAGTGGAC AACAGGGCCA TCCAGTAAC CTGCATCACA TCATGACTAC
751 AAATGTGCAA ATGTCTATCA TCCCGAGCAA TGCTCCTGGG CCCCCTCTTC
801 ACATTGGAGC TTCTCATTTA CCTCGAGGTG CAGCTGCTGC TGCTGTGATG
851 TCCAGTTCTA AAGTAACCAC AGTCTGAGG CCGACCTCAC AGCTGCCAAA
901 TGCTGCTACT GCTCAGCCAG CAGTACAGCA CATCATTCAC CAACCAATCC
951 AGTCTCGGCC ACCTGTGACC ACCTCCAATG CCATCCCTCC TGCTGTGGTA
1001 GCAACTGTCT CAGCCACCAG AGCTCAGTCT CCAGTCATCA CTACGACAGC
1051 GCGCGATGCT ACTGATTGAG CACTTAGTAG GCCAACCTTG TCTATCCAGC
1101 ATCTCTCATC TGCAGCAATC AGTATTGAGC GTCCTGCCCA GTCACGAGAT
1151 GTCACACAAA GAATCACTAC ACCATCTCAC CCTGCATTAG GGACGCCAAA
1201 ACAGCAGCTT CATACAATGG CTCAGAAAAC AATCTTCAGT ACTGGCAGCG
1251 CAGTGGCTGC AGCCACAGTA GCACCTATTT TGGCAACCAA CACCAATCCT
1301 TCAGCGACCA CAGCTGGATC TGTGTACAC ACCCAAGCTC CCACAAGTAC
1351 CATTTGTACC ATGACAGTAC CCTCCCATTC CTCCCATGCT ACTGTGTGTA
1401 CCACCTCAAA CATCCAGTCC GCCAAGGTGG TGCCCGAGCA GATCAGCCAC
1451 ACTTCTCCTC GGATCCAGCC AGACTACCTT GCCGAGAGGA GTAGCCTGAT
1501 TCCCATCTCC GGACATCGGG CCTCTCCCAA TCCTGTGGCC ATGGAAACCC
1551 GAAGTGACAA CAGACCGTCT GTTCCCGTTC AGTTCCAATA TTTTGTGCCA
1601 ACTTACCCCC CTCTGCATA CCCACTGGCG GCACATACCT ACACCCCAAT
1651 CACCAAGTCC GTGTCCACTA TCCGACAGTA TCCAGTTTCA GCTCAGGCTC
1701 CAAACTCTGC CATCACAGCT CAGACTGGTG TTGGGTAGC GTCTACCGTC
1751 CACCTAAACC CCATGCAGTT GATGACAGTG GATGCATCGC ATGCTCGACA
1801 TATTCAAGGG ATCCAGCCAG CACCCATCAG TACCCAGGGT ATCCAGCCGG
1851 CCCCATTGG GACCCAGGG ATACAGCCTG CACCATTGG CACACAGGGA
1901 ATTCACTCAG CAACCCCAAT CAACACAAA GGGCTTCAGC CTGCACCTAT
1951 GGTACTCAG CAGCCTCAGC CTGAAGGAAA GACTTCAGCA GTGGTGTGG
2001 CAGATGGAGC CACAATTGTG GCCAACCCTA TTAGCAATCC ATTCAGTGGT
2051 GCTCCAGCAG CAACAACCGT GGTGCAGACC CACAGCCAGA GTGCTAGCAC
2101 CAACGCTCCC GCCCAGGGCT CATCGCCACG GCCAAGCATA CTCCGGAAGA
2151 AACCTGCCAC AGATGGTGCC AAACCCAAGT CTGAAATCCA CGTGTCTATG
2201 GCCACTCCGG TCACTGTGTC CATGGAGACT GTATCCAATC AAAATAATGA
2251 TCAGCCTACC ATTGCCGTCC CTCCAAGTGC CCAGCAGCCC CCACCGACCA
2301 TTCCAATAT GATTGCAGCA GCCAGTCCCC CGTCACAACC AGCCGTGGCC
2351 CTTTCAACCA TTCTGGAGC GGTCCCATC ACTCCACCCA TCACCAACAT
2401 TGCAAGTGCA CCACCTCCAT CAGTCACTGT GGGTGGCAGT CTTTCTCCG
2451 TCTTGGGCCC TCCGTTCTCT GAAATTAAG TGAAGAAGA AGTAGAACCA
2501 ATGGATATCA TGAGGCCAGT TTCTGCAGTT CCTCACTGG CTACCAACAC
2551 TGTGTCTCCA TCTCTGCAT TGCTGGCAAA CAACTTGTCC ATGCCTACAA
2601 GTGACCTACC ACCTGGTGCC TCCCAAGGA AAAAGCCTCG AAAGCAACAG
2651 CATGTGATCT CAACAGAAGA AGGTGACATG ATGGAGACAA ACAGCACTGA
2701 TGATGAGAAG TCCACTGCCA AGAGTCTTCT GGTGAAGGCT GAGAAGCGCA
```

```

2751 AGTCTCCTCC CAAGGAGTAT ATTGATGAGG AAGGTGTGAG ATATGTCCCA
2801 GTGCGTCCAA GACCCCCCAT TACTTTGCTT CGTCACATC GGAACCCCTG
2851 GAAAGCTGCT TACCACCACT TTCAGAGGTA CAGTGACGTC CGGGTCAAAG
2901 AGGAGAAGAA AGCTATGCTG CAGGAAATAG CTAATCAGAA AGGAGTATCC
2951 TGTCTGTGCTC AAGGCTGGAA AGTCCACCTC TGTGCTGCCC AGTTACTACA
3001 GCTGACGAAT CTAGAACATG ATGTCTATGA AAGACTTACT AACCTGCAGG
3051 AAGGGATTAT CCCAAAGAAA AAAGCAGCAA CAGATGATGA TCTCCACCGA
3101 ATAAACGAAC TGATACAGGG AAATATGCAG AGGTGTAAAC TTGTGATGGA
3151 TCAATCAGT GAAGCCAGAG ACTCCATGCT TAAGGTTTAA GATCATAAAG
3201 ACCGTGCTCT GAAGCTGCTT AACAAAGAAC GGACTGTCAA AAAAGTGTCC
3251 AAATTGAAGC GAAAGGAAAA AGTCTAGACC CAGAACATC AGGAGATTGG
3301 AAGCAAATTT ATGAAGAATG ATGGTGGGGG TGGGGGGAGG GTTTTGGTTT
3351 TTTCCAAAGT GGAACATTGA AATAAAGGAA GTGTTCTTAA GTTCCCCTGT
3401 GAAAGCAGAG GAACCCATGA CATCCAAGG CGTGAAAGGA TCAGAGCTGA
3451 CTGGACATAG TGAGCTGCCT TCTTGCGTTC GGGTGCACCC CTGTTAAACC
3501 TGATCTGTGT CATAAGTGAC TCCGGATGCA TCAGTGTCCA CCAGTTGGAA
3551 GCAATGACAA GGATGGCTGG CTGGTGTGTT TCAGCCTTCC GGTTTATAGA
3601 CTGTATTAT CTAGTGGATT CCTGCAGGCC CCATACTGAG CCTGGACTGA
3651 AAGTATCCAC TCGGACCATC TGTATCTCT CTACACTGAA AATAAACCTT
3701 CTTCCACCCA CCCCATTCGG TTCTTCTGCC TGACCTTCAA ATGCCCATGT
3751 TGCCCTTTTA CAGCAGTGCC ACGGCACCAA GCGAGCTGCC ACATCTCACA
3801 CTCTAAAGGG TTTGAACTAT TAGTCTTGT CATTTTTAA AAAAAACCAT
3851 TCCCAAGTGA AATTGTTATA TCGTCTGTCT TCGTGTGTCT AGAACTGGGT
3901 TTTTGTGGAG GTTCAGAGCA GGCAACACCA TAAGTTGCTC TCAGATCCTT
3951 GTTCTGAAGT ACATTCTTGG TTATCTGTAC TTCTGTAGCT GGTGTGATGC
4001 TGTTAATTGT ATGTACCACA CATCTCCAGA CGTTAATAAA GGACTCAAAG
4051 AGGTTTTTGT AAAAAAAAAA AAAAAAAAAA AA

```

## BLAST Results

No BLAST result

## Medline entries

No Medline entry

## Peptide information for frame 2

ORF from 131 bp to 3274 bp; peptide length: 1048  
 Category: similarity to known protein

```

1 MGPPRHPQAG EIEAGGAGGG RRLQVEMSSQ QPRLGAPST GLSQAPSQIA
51 NSGSAGLINP AATVNDESGR DSEVSAREHM SSSSSLSQRE EKQEPVVVRP
101 YPQVQMLSTH HAVASATPVA VTAPPAHLTP AVPLSFSEGL MKPPFKPTMP
151 SRPIAPAPPS TSLSPKVPVG QVTVTMESSI PQASAI PVAT ISGQGHPSN
201 LHHIMTTNVQ MSIIRSNAPG PPLHIGASHL PRGAAAAAVM SSSKVTTVLR
251 PTSQLPNAAT AQPAVQHIIH QPIQSRPPVT TSNAIPPAVV ATVSATRAQS
301 PVITTTAAHA TDSALSRPTL SIQHPPSAAI SIORPAQSRD VTTRITLPSH
351 PALGTPKQQL HTMAQRTIFS TGTTPVAAATV APILATNTIP SATTAGSVSH
401 TQAPTSTIVT MTVPSSHSHA TAVTTSNIPV AKVVPPQIITH TSPRIQPDYP
451 AERSSLIPIS GHRASPNPVA METRSDNRPS VPVQFYFLP TYPPSAYPLA
501 AHYTPITISS VSTIRQYPVS AQAPNSAITA QTGVGVASTV HLNPMQLMTV
551 DASHARHIQG IQPAPISTQG IQPAPIGTPG IQPAPLGTQG IHSATPINTQ
601 GLQAPPMGTQ QPQPEGKTSV VVLADGATIV ANPISNPFSA APAATTVVQT
651 HSQASTNAP AQGSSPRPSI LRKKPATDGA KPKSEIHVSM ATPVTVMET
701 VSNQNNDOPT IAVPPTAQPF PPTIPTMIAA ASPPSQPAVA LSTIPGAVPI
751 TPPITTTAAA PPPSVTVGGS LSSVLGPPVP EIKVKEEVEP MDIMRPVSAV
801 PPLATNTVSP SLALLANNLS MPTSDLPPGA SPRKKPRKQQ HVISTEEGDM
851 METNSTDDEK STAKSLLVKA EKRKSPPEY IDEEGVRYVP VRPRPPITLL
901 RHYRNPWKAA YHHFQRYSDV RVKEKKAML QEIANQKGVs CRAQGWKVLH
951 CAAQLQLTN LEHDVYERLT NLQEGIIKK KAATDDDLHR INELIQGNMQ
1001 RCKLVMDQIS EARDSMLKVL DHKDRVLKLL NKNGTVKKVS KLKRKEKV

```

## BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphtes3\_2a11, frame 2

SWISSPROT:MUC2\_HUMAN MUCIN 2 PRECURSOR (INTESTINAL MUCIN 2)., N = 1,  
 Score = 334, P = 2.4e-25

PIR:A43932 mucin 2 precursor, intestinal - human (fragments), N = 1,  
Score = 321, P = 3.2e-24

TREMBL:D88440\_1 product: "high molecular mass nuclear antigen"; Gallus  
gallus mRNA for high molecular mass nuclear antigen, partial cds., N =  
1, Score = 312, P = 8.3e-24

PIR:S48478 glucan 1,4-alpha-glucosidase (EC 3.2.1.3) - yeast  
(Saccharomyces cerevisiae), N = 1, Score = 300, P = 2.1e-22

>SWISSPROT:MUC2\_HUMAN MUCIN 2 PRECURSOR (INTESTINAL MUCIN 2).  
Length = 5,179

## HSPs:

Score = 334 (50.1 bits), Expect = 2.4e-25, P = 2.4e-25  
Identities = 184/770 (23%), Positives = 263/770 (34%)

Query: 96 VVVRYPQVQMLSTHHAVASATPVAVTAPPAHL-TPAVPLSFSEGLMKPPKPTMPSRPI 154  
V P P T + + T V T P TP + + P P PT P  
Sbjct: 3471 VTPTPTGTGTPTTTPIITTTTPTPTPTGTGTPTTTPIITTTTPTPTPTGTGTPT 3530

Query: 155 A-PAPPSTLSLPPKVP-GQVTVMESSIPOASAI PVATISGQQGHPSNLHHIMTTNVQMS 212  
P +T P P G T T + P T +G Q P+ TT V +  
Sbjct: 3531 TTPITTTTPTPTPTGTGTPTTTPIITTTTPTPTPTGTGT-PTTPIITTTTPTPT 3589

Query: 213 IIRSNAPGP---PLHIGASHLPRGAAAAA-VMSSSKVTVLRPTSQLPNAATAQPAVQHI 268  
+ P P+ + P +++ +TT T T P I  
Sbjct: 3590 PTPTGTGTPTTTPIITTTTPTPTPTGTGTPTTTPIITTTTPTPTPTGTGTPTTTPI 3649

Query: 269 IHQIQSRPPVTSNAIPPAVVATVSATRAQSPVITTTAAHATDSALSRPTLSIQHPPSA 328  
+ P T P T + T +P T T T + T++ P  
Sbjct: 3650 TTTTPTPTPTGTGTPTTTPIITTTTPTPTPTGTGTPTTTPIITTTTPTPTPTPT 3706

Query: 329 AISIQRPQRSRDVTRITLPSHPALGTFKQQLHTMAQKT-IFSTGTPVAAAT--VAPILA 385  
Q P + TT P+ GT + T + T TP T PI  
Sbjct: 3707 PTGTGTPTTTPIITTTTPTPTPTGTGTPTTTPIITTTTPTPTPTGTGTPTTTPIIT 3766

Query: 386 TNTI-PSATTAGSVSHTQAPTSTIVTMT-VPSHSSHATAVTSNIPVAKVVPQQIHTSP 443  
T T+ P+ T G+ + T P +T T+T P+ + T TT V P T T  
Sbjct: 3767 TTTTPTPTPTGTGTPTTTPIITTTTPTPTPTGTGTPTTTPIITTTTPTPTPTPTPTGT 3825

Query: 444 RIQPDYPAERSSLIPISGHRASPNPVMETRSDNRPSVPVQFQYFL-PTYPPSAYPLAAH 502  
+ P ++ + +P P +T + + P+ + PT P+  
Sbjct: 3826 QTPTTTPIITTTTPTPTPTPTGTGTPTTTPIITTTTPTPTPTPTGTGTPTPTPTGT 3874

Query: 503 TYTPITSSVS-TIRQYPVSAQAPNSA-ITAGTGVGVASTVHLNPMQLMTVDASHARHIQG 560  
T TPIT++ + T P Q P + IT T V T Q T  
Sbjct: 3875 TTTTPTTTTPTPTPTGTGTPTTTPIITTTTPTPTPTPTPTPTPTPTPTPTPTPTPTPT 3932

Query: 561 IQPAPISTQGIQAPIGTPGI---QPAPLGTQGIHSATPINTQGL---QPAPMGTOQPQ- 613  
P P TQ PI T P P GTQ + TPI T P P GTQ P  
Sbjct: 3933 PTPTPTGTGTPTTTPIITTTTPTPTPTGTGT-PTTTPIITTTTPTPTPTGTGTPTT 3991

Query: 614 -PEGKTSAVVLADGATIVANPISNPFSAAPAAT-TVVQTHSQSASTNAPAQGSSPRPSIL 671  
P T+ V T P + P + T T T +Q+ +T ++ P+  
Sbjct: 3992 TPITTTTPTPTPTGTGTPTTTPIITTTTPTPTPTGTGTPTTTPIITTTTPTPTPTPT 4051

Query: 672 RKKPATDGAKPKSEIHVSMATPVTVSMETVSNQNDQPTIAVP---PTAQPPPTIPTMI 728  
T P + TP +T + T P PT Q P T P  
Sbjct: 4052 PTGTGTPTTTPIITTTTPTPTPTGTGTPTTTPIITTTTPTPTPTGTGTPTTTPIIT 4111

Query: 729 AAASPPSQPAVALSTIPGAVPITPPITIAAAPPS----VTVGSLSSVLGP-FVPEI 782  
P+ T P PIT TT+ P P+ T + ++ + P P P  
Sbjct: 4112 TTTVPTPTPTGTGT-PTTTPIITTTTPTPTPTGTGTPTTTPIITTTTPTPTPTPTGT 4169

Query: 783 KVKEEVEPMDIMRPVSAVP-PLATNTVSPSLALLANLMSMPTSDLPPGASPRKKPRKQOH 841  
P+ V+ P P T T P+ A + TS+ PP +S + R  
Sbjct: 4170 TQTPTTTPIITTTTPTPTPTGTGTGPTHTSTAPIAELTNSPPESPSTPQTSRSTSS 4229

Query: 842 VISTEEGDMMET 853  
+ TE ++ T  
Sbjct: 4230 PL-TESTLLST 4240

Score = 328 (49.2 bits), Expect = 1.0e-24, P = 1.0e-24  
Identities = 180/745 (24%), Positives = 254/745 (34%)

Query: 96 VVVRYPQVQMLSTHHAVASATPVAVTAPPAHL-TPAVPLSFSEGLMKPPKPTMPSRPI 154  
V P P T + + T V T P TP + + P P PT P  
Sbjct: 3540 VTPTPTGTGTPTTTPIITTTTPTPTPTGTGTPTTTPIITTTTPTPTPTGTGTPT 3599

Score = 325 (48.8 bits), Expect = 2.2e-24, P = 2.2e-24  
Identities = 186/782 (23%), Positives = 261/782 (33%)

764

T TPIT++ + T P Q P + IT T V T Q T  
 Sbjct: 3898 TTTPIITTTTVPPTPTGTQTPTTTPIITTTTVPPTPT--TGTQTPTTPIITTTTVP 3955  
 Query: 561 IQPAPISTQGIQAPAGITPGI---QPAPLGTQGIHSATPINTQGL---QPAPMGTOQPQ- 613  
 P P TQ PI T P P GTQ + TPI T P P GTQ P  
 Sbjct: 3956 PTPTPTGTQTPTTTPIITTTTVPPTPTGTQ-TPTTPIITTTTVPPTPTGTQTPTT 4014  
 Query: 614 -PEGKTSAVVLADGATIVANPISNPFSAAPAAT-TVVQTHSQSASTNAPAQGSSPRPSIL 671  
 P T+ V T P + P + T T T +Q+ +T ++ P+  
 Sbjct: 4015 TPITTTTVPPTPTGTQTPTTTPIITTTTVPPTPTGTQTPTTTPIITTTTVPPTPT 4074  
 Query: 672 RKKPATDGAKPKSEIHVSMATPVTVMETVSNQNDQPTIAVP---PTAQPPPTIPTMI 728  
 T P + TP +T + T P PT Q P T P  
 Sbjct: 4075 PTGTPTTTTPIITTTTVPPTPTGTQTPTTTPIITTTTVPPTPTGTQTPTTTPIIT 4134  
 Query: 729 AAAPSPQPAVALSTIPGAVPITPPITTTIAAAPSPSVTVGSSLSSVLGPPVPEIKVKEEV 788  
 P+ T P PIT TT P P+ T G+ + P I V  
 Sbjct: 4135 TTTVPTPTPTGTQT-PTTTPIT---TTTTVPTPTPT--GTQT---PTTTPITTTTVP 4184  
 Query: 789 EPMDIRPVSAPVPLATNTVSPSLALLANLSMPTSDLPSPGASPRKKPRKQOHVISTEEG 848  
 P PP T+T +P L +N P S P + P + + +  
 Sbjct: 4185 TPTPTPTGTQTGPPTHIST-APIAELTTSN-PPESSTPQTSRSTSSPLTESTTLLSTLP 4242  
 Query: 849 DMMETNSTDDEKSTAKSLLVKAERKSPP 877  
 +E ST + SPP  
 Sbjct: 4243 PAIEMTSTAPPSTPTAPTTTSGGHTLSPP 4271  
 Score = 324 (48.6 bits), Expect = 2.8e-24, P = 2.8e-24  
 Identities = 170/717 (23%), Positives = 248/717 (34%)  
 Query: 95 PVVVRFPYQVQMLSTHVASATP--VAVTAPPAHLTFAVPLSFSEGLMKPPKPTMPSR 152  
 P P P +T + +P T PP TP+ P++ + + P P+ P  
 Sbjct: 1401 PPTTTPSPPTTTTTLPTTTTPSPPTTTTTPPTTTPSPPTTTTTPPL-PTTTPSPPT 1459  
 Query: 153 PIAPAPSTLSLPPKVPQVVTMESSIPQASAIPTATISGQQGHPNLHHIMTINVQMS 212  
 PP+T PP T S + P T + P I +  
 Sbjct: 1460 TTTTTPPTTTPSPPTTTPSPPTTTPSPPTTTPPTTTPS---PMTTPTITPPASTTT 1516  
 Query: 213 IIRSNAPGPPPLHIGASHLPRGAAAAVMSSSKVTVLRTSQ--LPNAATAQPAVQHIIH 270  
 + + P PP + P S T + PTS LP T P  
 Sbjct: 1517 LPPTTTPSPPTTTTTPPP-----TTTTPSPPTTTPITPTTSTTLPTTTPSPPTTTT 1571  
 Query: 271 QPIQSRP-PVTSNAIPPAVVATVSA-TRAQSPVITTAHAATDSALSRPTLSIQHPPSA 328  
 P + P P TT+ PP + T T SP TTT + S PT + PP++  
 Sbjct: 1572 PPTTTPSPPTTTPSPPTITTTTTPPTTTPSPPTTTPPTTTPPTTTPSPPTTTPITPT 1631  
 Query: 329 AISIQRPQSRDVTTRITLPSHPALGTPKQQLHTMAQKTIFSTGTVAATVAPILATNT 388  
 ++ T T P P TP T I +T TP T + + T  
 Sbjct: 1632 TTTLPPTTTPSPPTTTTTP--PPTTTPSPPTTTPSPPTTTPPTTTPPTTTPSSPTTTP 1689  
 Query: 389 IPSATTAGSVSHTOAPTSTIVTMTVPSHSHATAV-TTSNIPVAKVVPQQITHTSPIRQP 447  
 P TT + S T P+S I T T PS ++ + TT P P T T + P  
 Sbjct: 1690 SPPTTMTTTPSPPTTTPSPPTTTTTPSSTTTPSPPTTMTTTPSPPTTTPSPPTTMTTLP 1749  
 Query: 448 DYPAERSSLIPIGHRASPNFVAMETRSDNRPSVPV-QFQYFLPTYPPSAY-P-----LA 500  
 + + P+ P T + P VP+ + +L + P+ + P L  
 Sbjct: 1750 TTTSSPLTTPLPSPITPPTTTPSPFTTTPPTPCVPLCNWTGWLDSGKPNFHKPGGDELI 1809  
 Query: 501 AHYYTPTITSSVSTIR--QYP-VSAQAPNSAITAQTVGV-VASTVHLNPMQLMTVDASHAR 556  
 P ++ + R YP V + VG + P ++ + A  
 Sbjct: 1810 GDVCGPGWAANISCRATMYPDVPIGQLGQTVVCDVSVGLICKNEDQKPGGVIPM-AFCLN 1868  
 Query: 557 HIQIQIAPAPISTQGIQAPAGITPGIQ-PAPLGTQGIHSATPINTQGLQAPAPMGTOQPQ-- 613  
 + +Q TQ P + T + P P T I + T + P P GTQ P  
 Sbjct: 1869 YEINVQCECVTQ---PTTMTTTTENPTPTTPTTPIITTTTVPPT---PTPTGTQTPTT 1922  
 Query: 614 PEGKTSAVVLADGATIVANPISNPFSAAPAAT-TVVQTHSQSASTNAPAQGSSPRPSILR 672  
 P T+ V T P + P + T T T +Q+ +T ++ P+  
 Sbjct: 1923 PIITTTTVPPTPTGTQTPTTTPIITTTTVPPTPTGTQTPTTTPIITTTTVPPTPTPT 1982  
 Query: 673 KKPATDGAKPKSEIHVSMATPVTVMETVSNQNDQPTIAVP---PTAQPPPTIPTMIA 729  
 T P + TP +T + T P PT Q P T P  
 Sbjct: 1983 TGTQTPTTTTPIITTTTVPPTPTGTQTPTTTPIITTTTVPPTPTGTQTPTTTPIITT 2042  
 Query: 730 AAAPSPQPAVALSTIPGAVPITPPITTTIAAAPSPSVTVGSSLSSVLGPPVPEIKVKEEVE 789  
 P+ T P PIT TT P P+ T G+ + P V  
 Sbjct: 2043 TTVTPTPTGTQT-PTTTPIT---TTTTVPTPTPT--GTQTPTTPIITTTTVPPTPT 2096  
 Query: 790 PMDIRPVSAPVPLATNTVSPS 811  
 P P + P T TV+P+  
 Sbjct: 2097 PTGTQTPTTT-PITTTTVPPTPT 2117

Score = 318 (47.7 bits), Expect = 1.2e-23, P = 1.2e-23  
Identities = 174/717 (24%), Positives = 243/717 (33%)

Query: 96 VVVRYPQVQMLSTHHAVASATPVAVTAPPAHL-TPAVPLSFSEGLMKPPPKPTMPSRPI 154  
V P P T + + T V T P TP + + P P PT P  
Sbjct: 2068 VTPTPTGTGTPTTTPIITTTTPTPTPTGTGTPTTTPIITTTTPTPTPTGTGTPT 2127

Query: 155 A-PAPPSTLSLPPKVP-GQVTVMESSIPQASAI PVATISGQOGHPSNLHHIMTTNVQMS 212  
P +T P P G T T + P T +G Q P+ TT V +  
Sbjct: 2128 TTPITTTTPTPTPTGTGTPTTTPIITTTTPTPTPTGTGT-PTTPIITTTTPTPT 2186

Query: 213 IIRSNAPGP---PLHIGASHLPRGAAAAA-VMSSSKVTTVLRPTSQLPNAATAQPAVQHI 268  
+ P P+ + P +++ +TT T T P I  
Sbjct: 2187 PTPTGTGTPTTTPIITTTTPTPTPTGTGTPTTTPIITTTTPTPTPTGTGTPTTTPI 2246

Query: 269 IHQIQSRPPVTTSSNAIPPAVVATVSATRAQSPVITTTAAHATDSALSRLPSIQHPFSA 328  
+ P T P T + T +P T T T + T++ P  
Sbjct: 2247 TTTTPTPTPTGTGTPTTTPIITTTTPTPTPTGTGTPTTTPIITTTTPTPTPT 2303

Query: 329 AISIQRPQSRDVTTRITLPSHPALGTPKQQLHTMAQKT-IFSTGTPVAAAT--VAPILA 385  
Q P + TT P+ GT + T + T TP T PI  
Sbjct: 2304 PTGTGTPTTTPIITTTTPTPTPTGTGTPTTTPIITTTTPTPTPTGTGTPTTTPIIT 2363

Query: 386 TNTI-PSATTAGSVSHQAPTSTIVTMT-VPSHSSHATAVTTSNIPVAKVVPQIHTTSP 443  
T T+ P+ T G+ + T P +T T+ T TT V P T T  
Sbjct: 2364 TTTTPTPTPTGTGTPTTTPIITTTTPTPTPTGTGTPTTTPIITTTTPTPTPTPTGT 2422

Query: 444 RIQPDYPAERSSLIPIGHRASPNPVAMETRSNRPVSVQFQYFL-PTYPPSAYPLAAH 502  
+ P ++ + +P P +T + +P+ + PT P+  
Sbjct: 2423 QTPTTTPIITTTTPTPTPTPTGTGTPTTTPIITTTTPTPTPTPTGTGT--TQTP 2471

Query: 503 TYTPITSSVS-TIRQYPVSAQAPNSA-ITAQTGVGVASTVHLNPMQLMTVDASHARHIQ 560  
T TPIT++ + T P Q P + IT T V T Q T  
Sbjct: 2472 TTTTPTTTTPTPTPTGTGTPTTTPIITTTTPTPTPTPTPT--TGQTPTTTPIITTTTPT 2529

Query: 561 IQPAPISQGIQAPIGTPTGI--QPAPLGTQGIHSATPINTQGL--QPAPMGTTQFPQ- 613  
P P TQ PI T P P GTQ + TPI T P P GTQ P  
Sbjct: 2530 PTPTPTGTGTPTTTPIITTTTPTPTPTGTGT-TPTTPIITTTTPTPTPTGTGTPTT 2588

Query: 614 -PEGKTSAVVLADGATIVANPISNPFSAAPAAAT-TVVQTHSQSASTNAPAQSSPRPSIL 671  
P T+ V T P + P + T T T +Q+ +T ++ P+  
Sbjct: 2589 TPTTTTPTPTPTGTGTPTTTPIITTTTPTPTPTGTGTPTTTPIITTTTPTPTPT 2648

Query: 672 RKKPATDGAKEIHSVSMATPVTVSMETVSNQNDQPTIAPV---PTAQPPPTIPTMI 728  
T P + TP +T + T P PT Q P T P  
Sbjct: 2649 PTGTGTPTTTPIITTTTPTPTPTGTGTPTTTPIITTTTPTPTPTGTGTPTTTPIIT 2708

Query: 729 AAASPPSQPAVALSTIPGAVPITPITTTIAAAPPSVTVGGSLSSVLGPPVPEIKVKEEV 788  
P+ T P PIT TT P P+ T G+ + P V  
Sbjct: 2709 TTTTPTPTPTGTGT-PTTTPIT---TTTTPTPTPT--GTGTPTTTPIITTTTPTPT 2762

Query: 789 EPMDIMRPVSAVPLATNTVSPS 811  
P P + P T TV+P+  
Sbjct: 2763 TPTGTGTPTTT-PITTTTPTPT 2784

Score = 318 (47.7 bits), Expect = 1.2e-23, P = 1.2e-23  
Identities = 174/717 (24%), Positives = 243/717 (33%)

Query: 96 VVVRYPQVQMLSTHHAVASATPVAVTAPPAHL-TPAVPLSFSEGLMKPPPKPTMPSRPI 154  
V P P T + + T V T P TP + + P P PT P  
Sbjct: 2206 VTPTPTGTGTPTTTPIITTTTPTPTPTGTGTPTTTPIITTTTPTPTPTGTGTPT 2265

Query: 155 A-PAPPSTLSLPPKVP-GQVTVMESSIPQASAI PVATISGQOGHPSNLHHIMTTNVQMS 212  
P +T P P G T T + P T +G Q P+ TT V +  
Sbjct: 2266 TTPITTTTPTPTPTGTGTPTTTPIITTTTPTPTPTGTGT-PTTPIITTTTPTPT 2324

Query: 213 IIRSNAPGP---PLHIGASHLPRGAAAAA-VMSSSKVTTVLRPTSQLPNAATAQPAVQHI 268  
+ P P+ + P +++ +TT T T P I  
Sbjct: 2325 PTPTGTGTPTTTPIITTTTPTPTPTGTGTPTTTPIITTTTPTPTPTGTGTPTTTPI 2384

Query: 269 IHQIQSRPPVTTSSNAIPPAVVATVSATRAQSPVITTTAAHATDSALSRLPSIQHPFSA 328  
+ P T P T + T +P T T T + T++ P  
Sbjct: 2385 TTTTPTPTPTGTGTPTTTPIITTTTPTPTPTGTGTPTTTPIITTTTPTPTPT 2441

Query: 329 AISIQRPQSRDVTTRITLPSHPALGTPKQQLHTMAQKT-IFSTGTPVAAAT--VAPILA 385  
Q P + TT P+ GT + T + T TP T PI  
Sbjct: 2442 PTGTGTPTTTPIITTTTPTPTPTGTGTPTTTPIITTTTPTPTPTGTGTPTTTPIIT 2501

Query: 386 TNTI-PSATTAGSVSHQAPTSTIVTMT-VPSHSSHATAVTTSNIPVAKVVPQIHTTSP 443  
T T+ P+ T G+ + T P +T T+ T P+ + T TT V P T T  
Sbjct: 2442 PTGTGTPTTTPIITTTTPTPTPTGTGTPTTTPIITTTTPTPTPTGTGTPTTTPIIT 2501

Sbjct: 2502 TTTVTPTPTGTGTPTTTTPIITTTTPTPTPTGTGTPTTTTPIITTTTPTPTPTPTGTGT 2560

Query: 444 RIQPDYPAERSSLIPISGHRASPNPVAMETRSDNRPSVPVQFQYFL-PTYPPSAYPLAAH 502  
+ P ++ + +P P +T + + P+ + PT P+

Sbjct: 2561 QTPTTTPITTTTPTVT-----PTPTPTGTGTPT-----TTPITTTTPTPTPTPTGT--TQTP 2609

Query: 503 TYTPITSSVS-TIRQYVSAQAPNSA-ITAQTGVGVASTVHLNPMQLMTVDASHARHIQ 560  
T TPIT++ + T P Q P + IT T V T Q T

Sbjct: 2610 TTTPIITTTTPTPTPTPTGTGTPTTTTPIITTTTPTPTPTPT--TGTQTPTTPTITTTTPT 2667

Query: 561 IQPAPISTQGIQAPIGTPTGI---QPAPLGTQGIHSATPINTQGL---QPAPMGTQQPQ- 613  
P P TQ PI T P P GTQ + TPI T P P GTQ P

Sbjct: 2668 PTPTPTGTGTPTTTTPIITTTTPTPTPTGTGT-TPTTTPITTTTPTPTPTGTGTPTT 2726

Query: 614 -PEGKTSAVVLADGATIVANPISNPFSAAPAAT-TVVQTHSQSASTNAPAQSSPRPSIL 671  
P T+ V T P + P + T T T +Q+ +T ++ P+

Sbjct: 2727 TPITTTTPTPTPTGTGTPTTTTPIITTTTPTPTPTGTGTPTTTTPIITTTTPTPTPT 2786

Query: 672 RKKPATDGAKPKSEIHVSMATPVTVMETVSNQNDQPTIAVP---PTAQPPPTIPTMI 728  
T P + TP +T + T P PT Q P T P

Sbjct: 2787 PTGTGTPTTPTITTTTPTPTPTGTGTPTTPTITTTTPTPTPTGTGTPTTPTITT 2846

Query: 729 AAASPPSQPAVALSTIPGAVPITPPITTTIAAAPPSVTVGGSLSVLGPPVPEIKVKEEV 788  
P+ T P PIT TT P P+ T G+ + P V

Sbjct: 2847 TTTVTPTPTPTGTGT-PTTTPIT---TTTTVTPTPTPT--GTQTPTTPTITTTTPTPT 2900

Query: 789 EPMDIMRPVSAPPLATNTVSPS 811  
P P + P T V+P+

Sbjct: 2901 TPTGTGTPTTT-PITTTTPTPT 2922

Score = 318 (47.7 bits), Expect = 1.2e-23, P = 1.2e-23  
Identities = 174/717 (24%), Positives = 243/717 (33%)

Query: 96 VVVRYPQVQMLSTHHAVASATPVAVTAPPAHL-TPAVPLSFSEGLMKPPPKPTMPSRPI 154  
V P P T + + T V T P TP + + P P PT P

Sbjct: 2321 VTPTPTPTGTGTPTTTTPIITTTTPTPTPTGTGTPTTTTPIITTTTPTPTPTGTGTPT 2380

Query: 155 A-PAPPSTLSLPPKVP-GQVVTMESSIPQASAI PVATISGQQGHPSNLHHIMTTNVOMS 212  
P +T P P G T T + P T +G Q P+ TT V +

Sbjct: 2381 TTPITTTTPTPTPTGTGTPTTTTPIITTTTPTPTPTGTGT-PTTTPITTTTPTPT 2439

Query: 213 IIRSNAPGP---PLHIGASHLPRGAAAAA-VMSSSKVTVLRPTSQLENAATAQPAVQHI 268  
+ P P+ + P +++ +TT T T P I

Sbjct: 2440 PTPTGTGTPTTTTPIITTTTPTPTPTGTGTPTTTTPIITTTTPTPTPTGTGTPTTTPI 2499

Query: 269 IHQPIQSRPFVTTNSAIPPAVVATVSATRAQSPVITTTAAHATDSALSRPTLSIQHPSPA 328  
+ P T P T + T +P T T T +T++ P

Sbjct: 2500 TTTTPTPTPTGTGTPTTTTPIITTTTPTPTPTGTGTPTTTTPIITTTTPTPTPT 2556

Query: 329 AISIQRPQSRDVTTRITLPSHPALGTPKQQLHTMAQKT-IFSTGTPVAAAT--VAPILA 385  
Q P + TT P+ GT + T + T TP T PI

Sbjct: 2557 PTGTQTPTTPTITTTTPTPTPTGTGTPTTPTITTTTPTPTPTGTGTPTTPTITT 2616

Query: 386 TNTI-PSATTAGSVSHTQAPTSTIVTMT-VPSHSSHATAVTTSNIPVAKVVPQIHTSP 443  
T T+ P+ T G+ + T P +T T+T P+ + T TT V P T T

Sbjct: 2617 TTTVTPTPTGTGTPTTTTPIITTTTPTPTPTGTGTPTTTTPIITTTTPTPTPTPTGTGT 2675

Query: 444 RIQPDYPAERSSLIPISGHRASPNPVAMETRSDNRPSVPVQFQYFL-PTYPPSAYPLAAH 502  
+ P ++ + +P P +T + + P+ + PT P+

Sbjct: 2676 QTPTTTPITTTTPTVT-----PTPTPTGTGTPT-----TTPITTTTPTPTPTPTGT--TQTP 2724

Query: 503 TYTPITSSVS-TIRQYVSAQAPNSA-ITAQTGVGVASTVHLNPMQLMTVDASHARHIQ 560  
T TPIT++ + T P Q P + IT T V T Q T

Sbjct: 2725 TTTPIITTTTPTPTPTGTGTPTTTTPIITTTTPTPTPTPT--TGTQTPTTPTITTTTPT 2782

Query: 561 IQPAPISTQGIQAPIGTPTGI---QPAPLGTQGIHSATPINTQGL---QPAPMGTQQPQ- 613  
P P TQ PI T P P GTQ + TPI T P P GTQ P

Sbjct: 2783 PTPTPTGTGTPTTTTPIITTTTPTPTPTGTGT-TPTTTPITTTTPTPTPTGTGTPTT 2841

Query: 614 -PEGKTSAVVLADGATIVANPISNPFSAAPAAT-TVVQTHSQSASTNAPAQSSPRPSIL 671  
P T+ V T P + P + T T T +Q+ +T ++ P+

Sbjct: 2842 TPITTTTPTPTPTGTGTPTTTTPIITTTTPTPTPTGTGTPTTTTPIITTTTPTPTPT 2901

Query: 672 RKKPATDGAKPKSEIHVSMATPVTVMETVSNQNDQPTIAVP---PTAQPPPTIPTMI 728  
T P + TP +T + T P PT Q P T P

Sbjct: 2902 PTGTQTPTTPTITTTTPTPTPTGTGTPTTTTPIITTTTPTPTPTGTGTPTTPTITT 2961

Query: 729 AAASPPSQPAVALSTIPGAVPITPPITTTIAAAPPSVTVGGSLSVLGPPVPEIKVKEEV 788  
P+ T P PIT TT P P+ T G+ + P V

Sbjct: 2962 TTTVTPTPTPTGTGT-PTTTPIT---TTTTVTPTPTPT--GTQTPTTPTITTTTPTPT 3015

Query: 789 EPMDIMRPVSAPVPLATNTVSPS 811  
 P + P T TV+P+  
 Sbjct: 3016 TPTGTQTPTTT-PITTTTIVTPT 3037

Score = 318 (47.7 bits), Expect = 1.2e-23, P = 1.2e-23  
 Identities = 174/717 (24%), Positives = 243/717 (33%)

Query: 96 VVVRYPYQVQMLSTHHAVASATPVAVTAPPAHL-TPAVPLSFSEGLMKPPPKPTMPSRPI 154  
 V P P T + + TV T P TP + + P P PT P  
 Sbjct: 2390 VTPTPTPTGTQTPTTTPIITTTTIVTPTPTGTQTPTTTPIITTTTIVTPTPTGTQTP 2449

Query: 155 A-PAPPSTLSLPPKVP-GQVVTMESSIPQASAI PVATISGQQGHPSNLHHIMTTNVQMS 212  
 P +T P P G T T + P T +G Q P+ TT V +  
 Sbjct: 2450 TTPITTTTIVTPTPTGTQTPTTTPIITTTTIVTPTPTGTQTP-TTTPITTTTIVTPT 2508

Query: 213 IIRSNAPGP---PLHIGASHLPRGAAAAA-VMSSSKVTVLRPTSQLPNAATAQPAVQHI 268  
 + P P+ + P +++ +TT T T P I  
 Sbjct: 2509 PTPTGTQTPTTTPIITTTTIVTPTPTGTQTPTTTPIITTTTIVTPTPTGTQTPTTTPI 2568

Query: 269 IHQPIQSRPPVTTNSAIPPAVVATVSATRAQSPVITTTAAHATDSALSRPTLSIQHPPSA 328  
 + P T P T + T +P T T T + T++ P  
 Sbjct: 2569 TTTTIVTPTPTGTQTPTTTPIITTTTIVTPTPTGTQTPTTTPIITTTTIVT---PTPT 2625

Query: 329 AISIQRPAQSRDVTTRITLPSHPALGTPKQQLHTMAQKT-IFSTGTPVAAAT--VAPILA 385  
 Q P + TT P+ GT + T + T TP T PI  
 Sbjct: 2626 PTGTQTPTTTPIITTTTIVTPTPTGTQTPTTTPIITTTTIVTPTPTGTQTPTTTPIIT 2685

Query: 386 TNII-PSATTAGSVSHTOAPTSTIVTMT-VPSHSSHATAVTTNSIPVAKVVPQQITHTSP 443  
 T T+ P+ T G+ + T P +T T+T P+ + T TT V P T T  
 Sbjct: 2686 TTTVTPTPTPTGTQTPTTTPIITTTTIVTPTPTGTQTPTTTPIITTTTIVTPTP-TPTGT 2744

Query: 444 RIQPDYPAERSSLIPIGHRASPNPVMETRSDNRPSVPVQFYFL-PTYPPSAYPLAAH 502  
 + P ++ + +P P +T + + P+ + PT P+  
 Sbjct: 2745 QTPTTPIITTTTIVT---PTPTPTGTQTP---TTPITTTTIVTPTPTPTG--TQTP 2793

Query: 503 TYTPTSSVS-TIRQYVPSAQAPNSA-ITAQTGVGVASTVHLNPMQLMTVDASHARHIQG 560  
 T TPIT++ + T P Q P + IT T V T Q T  
 Sbjct: 2794 TTTPIITTTTIVTPTPTPTGTQTPTTTPIITTTTIVTPTPTP--TGTQTPTTTPIITTTTIVT 2851

Query: 561 IQPAPISTQGIQAPIGTPIGI---QPAPLGTQGIHSATPINTQGL---QPAFMGTQQPQ- 613  
 P P TQ PI T P P GTQ + TPI T P P GTQ P  
 Sbjct: 2852 PTPTPTGTQTPTTTPIITTTTIVTPTPTPTGTQ-TPTTPIITTTTIVTPTPTPTGTQTP 2910

Query: 614 -PEGKTSAVVLADGATIVANPISNPFSAAPAAT-TVVQTHSQSASTNAPAQGSSPRPSIL 671  
 P T+ V T P + P + T T T +Q+ +T ++ P+  
 Sbjct: 2911 TPITTTTIVTPTPTPTGTQTPTTTPIITTTTIVTPTPTGTQTPTTTPIITTTTIVTPTPT 2970

Query: 672 RKKPATDGAQPKSEIHVSMATPVTVSMETVSNQNDQPTIAVP---PTAQPPPTIPTMI 728  
 T P + TP +T + T P PT Q P T P  
 Sbjct: 2971 PTGTQTPTTTPIITTTTIVTPTPTGTQTPTTTPIITTTTIVTPTPTPTGTQTPTTTPIIT 3030

Query: 729 AAASPPSQPAVALSTIPGAVPITPITTTIAAAPPSVTVGGSLSVLGPPVPEIKVKEEV 788  
 P+ T P PIT TT P P+ T G+ + P V  
 Sbjct: 3031 TTTVTPTPTPTGTQTP-TTTPIT---TTTTVTPTPTPT--GTQTPTTTPIITTTTIVTPT 3084

Query: 789 EPMDIMRPVSAPVPLATNTVSPS 811  
 P + P T TV+P+  
 Sbjct: 3085 TPTGTQTPTTT-PITTTTIVTPT 3106

Score = 318 (47.7 bits), Expect = 1.2e-23, P = 1.2e-23  
 Identities = 174/717 (24%), Positives = 243/717 (33%)

Query: 96 VVVRYPYQVQMLSTHHAVASATPVAVTAPPAHL-TPAVPLSFSEGLMKPPPKPTMPSRPI 154  
 V P P T + + TV T P TP + + P P PT P  
 Sbjct: 2459 VTPTPTPTGTQTPTTTPIITTTTIVTPTPTGTQTPTTTPIITTTTIVTPTPTGTQTP 2518

Query: 155 A-PAPPSTLSLPPKVP-GQVVTMESSIPQASAI PVATISGQQGHPSNLHHIMTTNVQMS 212  
 P +T P P G T T + P T +G Q P+ TT V +  
 Sbjct: 2519 TTPITTTTIVTPTPTGTQTPTTTPIITTTTIVTPTPTGTQTP-TTTPITTTTIVTPT 2577

Query: 213 IIRSNAPGP---PLHIGASHLPRGAAAAA-VMSSSKVTVLRPTSQLPNAATAQPAVQHI 268  
 + P P+ + P +++ +TT T T P I  
 Sbjct: 2578 PTPTGTQTPTTTPIITTTTIVTPTPTGTQTPTTTPIITTTTIVTPTPTGTQTPTTTPI 2637

Query: 269 IHQPIQSRPPVTTNSAIPPAVVATVSATRAQSPVITTTAAHATDSALSRPTLSIQHPPSA 328  
 + P T P T + T +P T T T + T++ P  
 Sbjct: 2638 TTTTIVTPTPTPTGTQTPTTTPIITTTTIVTPTPTGTQTPTTTPIITTTTIVT---PTPT 2694

Query: 329 AISIQRPAQSRDVTTRITLPSHPALGTPKQQLHTMAQKT-IFSTGTPVAAAT--VAPILA 385  
 Q P + TT P+ GT + T + T TP T PI  
 Sbjct: 2695 PTGTQTPTTTPIITTTTIVTPTPTGTQTPTTTPIITTTTIVTPTPTPTGTQTPTTTPIIT 2754



Query: 386 TNIT-PSATTAGSVSHTQAPTSTIVTMT-VPSHSSHATAVTTSNIPVAKVVPQQIHTHTSP 443  
T T+ P+ T G+ + T P +T T+T P+ + T TT V P T T  
Sbjct: 2755 TTTVTPTPTGTQTPTTTPITTTTTPPTPTGTQTPTTTPITTTTTPPTPTGT 2813

Query: 444 RIQPDYPAERSSLIPISGHRASPNVAMETRSNRPSPVQFQYFL-PTYPPSAYPLAAH 502  
+ P ++ + +P P +T + + P+ + PT P+  
Sbjct: 2814 QTPTTTPITTTTTPVT-----PTPTPTGTQTPT-----TTPITTTTTPPTPTGT--TQTP 2862

Query: 503 TYTPITSSVS-TIRQYPVSAQAPNSA-ITAQTGVGVASTVHLNPMQLMTVDASHARHIQG 560  
T TPIT++ + T P Q P + IT T V T Q T  
Sbjct: 2863 TTTPTTTTTPPTPTPTGTQTPTTTPITTTTTPPTPTPTGT--TGTQTPTTTPITTTTTPVT 2920

Query: 561 IQPAPISTQGIQAPIGTPTGI---QPAPLGTQGIHSATPINTQGL---QPAPMGTTQQPQ- 613  
P P TQ PI T P P GTQ + TPI T P P GTQ P  
Sbjct: 2921 PTPTPTGTQTPTTTPITTTTTPPTPTPTGTQ-TPTTTPITTTTTPPTPTPTGTQTPTT 2979

Query: 614 -PEGKTSAVVLADGATIVANPISNPFSAAPAAT-TVVQTHSQSASTNAPAQSSPRPSIL 671  
P T+ V T P + P + T T T +Q+ +T ++ P+  
Sbjct: 2980 TPITTTTTPPTPTPTGTQTPTTTPITTTTTPPTPTPTGTQTPTTTPITTTTTPPTPT 3039

Query: 672 RKKPATDGAKPKSEIHVSMATPVTVSMETVSNQNDQPTIAPV---PTAQPPPTIPTMI 728  
T P + TP +T + T P PT Q P T P  
Sbjct: 3040 PTGTQTPTTTPITTTTTPPTPTPTGTQTPTTTPITTTTTPPTPTPTGTQTPTTTPITT 3099

Query: 729 AAASPPSQPAVALSTIPGAVPITPPIITIAAAPPSVTVGGSLSVLGPPVPEIKVKEEV 788  
P+ T P PIT TT P P+ T G+ + P V  
Sbjct: 3100 TTTVTPTPTPTGTQT-PTTTPIT---TTTTVTPTPTPT--GTQTPTTTPITTTTTPPTPT 3153

Query: 789 EPMDIMRPVSAVPLATNTVSPS 811  
P P + P T TV+P+  
Sbjct: 3154 TPTGTQTPTTTPITTTTTPPTPT 3175

Score = 318 (47.7 bits), Expect = 1.2e-23, P = 1.2e-23  
Identities = 174/717 (24%), Positives = 243/717 (33%)

Query: 96 VVVRYPYQVQMLSTHHAVASATPVAVTAPPAHL-TPAVPLSFSEGLMKPPPKPTMPSRPI 154  
V P P T + + T V T P TP + + P P PT P  
Sbjct: 2528 VTPTPTPTGTQTPTTTPITTTTTPPTPTPTGTQTPTTTPITTTTTPPTPTPTGTQTPT 2587

Query: 155 A-PAPPSTLSLPPKVP-GQVTVMESSIPQASAI PVATISGQQGHPSNLHHIMTTNVQMS 212  
P +T P P G T T + P T +G Q P+ TT V +  
Sbjct: 2588 TTPITTTTTPPTPTPTGTQTPTTTPITTTTTPPTPTPTGTQT-PTTTPITTTTTPPTPT 2646

Query: 213 IIRSNAPGP---PLHIGASHLPRGAAAAA-VMSSSKVTTVLRPTSQLPNAATAQPAVQHI 268  
+ P P+ + P +++ +TT T T P I  
Sbjct: 2647 PTPTGTQTPTTTPITTTTTPPTPTPTGTQTPTTTPITTTTTPPTPTPTGTQTPTTTP 2706

Query: 269 IHQPIQSRPPVTTSNAIPPAVVATVSATRAQSPVITTTAAHATDSALSRLSLIQHPPSA 328  
+ P T P T + T +P T T T + T++ P  
Sbjct: 2707 TTTTTPPTPTPTGTQTPTTTPITTTTTPPTPTPTGTQTPTTTPITTTTTPVT---PTPT 2763

Query: 329 AISIQRPQSRDVTTRITLPSHPALGTPKQQLHTMAQKT-IFSTGTPVAAAT--VAPILA 385  
Q P + TT P+ GT + T + T TP T PI  
Sbjct: 2764 PTGTQTPTTTPITTTTTPPTPTPTGTQTPTTTPITTTTTPPTPTPTGTQTPTTTPITT 2823

Query: 386 TNIT-PSATTAGSVSHTQAPTSTIVTMT-VPSHSSHATAVTTSNIPVAKVVPQQIHTHTSP 443  
T T+ P+ T G+ + T P +T T+T P+ + T TT V P T T  
Sbjct: 2824 TTTVTPTPTPTGTQTPTTTPITTTTTPPTPTPTGTQTPTTTPITTTTTPPTPTPTGT 2882

Query: 444 RIQPDYPAERSSLIPISGHRASPNVAMETRSNRPSPVQFQYFL-PTYPPSAYPLAAH 502  
+ P ++ + +P P +T + + P+ + PT P+  
Sbjct: 2883 QTPTTTPITTTTTPVT-----PTPTPTGTQTPT-----TTPITTTTTPPTPTGT--TQTP 2931

Query: 503 TYTPITSSVS-TIRQYPVSAQAPNSA-ITAQTGVGVASTVHLNPMQLMTVDASHARHIQG 560  
T TPIT++ + T P Q P + IT T V T Q T  
Sbjct: 2932 TTTPTTTTTPPTPTPTGTQTPTTTPITTTTTPPTPTPTGT--TGTQTPTTTPITTTTTPVT 2989

Query: 561 IQPAPISTQGIQAPIGTPTGI---QPAPLGTQGIHSATPINTQGL---QPAPMGTTQQPQ- 613  
P P TQ PI T P P GTQ + TPI T P P GTQ P  
Sbjct: 2990 PTPTPTGTQTPTTTPITTTTTPPTPTPTGTQ-TPTTTPITTTTTPPTPTPTGTQTPTT 3048

Query: 614 -PEGKTSAVVLADGATIVANPISNPFSAAPAAT-TVVQTHSQSASTNAPAQSSPRPSIL 671  
P T+ V T P + P + T T T +Q+ +T ++ P+  
Sbjct: 3049 TPITTTTTPPTPTPTGTQTPTTTPITTTTTPPTPTPTGTQTPTTTPITTTTTPPTPT 3108

Query: 672 RKKPATDGAKPKSEIHVSMATPVTVSMETVSNQNDQPTIAPV---PTAQPPPTIPTMI 728  
T P + TP +T + T P PT Q P T P  
Sbjct: 3109 PTGTQTPTTTPITTTTTPPTPTPTGTQTPTTTPITTTTTPPTPTPTGTQTPTTTPITT 3168

Query: 729 AAASPPSQPAVALSTIPGAVPITPPIITIAAAPPSVTVGGSLSVLGPPVPEIKVKEEV 788

P+ T P PIT TT P P+ T G+ + P V  
Sbjct: 3169 TTTVTPTPTGTQT-PTTTPIT---TTTVPPTPTPT--GTQTPTTTPITTTTVPPTP 3222

Query: 789 EPMDIMRPVSAVPPLATNTVSPS 811  
P P+ P T TV+P+  
Sbjct: 3223 TPTGTQTPTT-PITTTTVPPT 3244

Score = 318 (47.7 bits), Expect = 1.2e-23, P = 1.2e-23  
Identities = 174/717 (24%), Positives = 243/717 (33%)

Query: 96 VVVRPYPQVQMLSTHHAVASATPVAVTAPPAHL-TPAVPLSFSEGLMKPPPKPTMPSRPI 154  
V P P T + + TV T P TP + + P P PT P  
Sbjct: 3080 VTPTPTPTGTQTPTTTPITTTTVPPTPTPTGTQTPTTTPITTTTVPPTPTPTGTQTPT 3139

Query: 155 A-PAPPSTLSLPPKVP-GQVVTMESSIPQASAI PVATISGQQGHPSNLHHIMTTNVQMS 212  
P +T P P G T T + P T +G Q P+ TT V +  
Sbjct: 3140 TTPITTTTVPPTPTPTGTQTPTTTPITTTTVPPTPTPTGTQT-PTTTPITTTTVPPT 3198

Query: 213 IIRSNAPGP---PLHIGASHLPRGAAAAA-VMSSSKVTVLRPTSQLPNAATAQPAVQHI 268  
+ P P+ + P +++ +TT T T P I  
Sbjct: 3199 PTPTGTQTPTTTPITTTTVPPTPTPTGTQTPTTTPITTTTVPPTPTPTGTQTPTTTP 3258

Query: 269 IHQIQSRPPVTTSNAI PPAVVATVSATRAQSPVITTTAAHATDSALSRPTLSIQHPPSA 328  
+ P T P T + T +P T T T + T++ P  
Sbjct: 3259 TTTTVPPTPTPTGTQTPTTTPITTTTVPPTPTPTGTQTPTTTPITTTTVP---PTPT 3315

Query: 329 AISIQRAQSRDVTTRITLPSHPALGTPKQQLHTMAQKT-IFSTGTPVAAT--VAPILA 385  
Q P + TT P+ GT + T + T TP T PI  
Sbjct: 3316 PTGTPTTTPITTTTVPPTPTPTGTQTPTTTPITTTTVPPTPTPTGTQTPTTTPITT 3375

Query: 386 TNTI-PSATTAGSVSHTQAPTSTIVTMT-VPSHSSHATAVTTSNIPVAKVVPQQITHTS 443  
T T+ P+ T G+ + T P +T T+T P+ + T TT V P T T  
Sbjct: 3376 TTTVPPTPTPTGTQTPTTTPITTTTVPPTPTPTGTQTPTTTPITTTTVPPTPTPTGT 3434

Query: 444 RIQPDYPAERSSLIPISGHRASPNVAMETRSDNRPSVPVQFYEL-PTYPPSAYPLAAH 502  
+ P ++ + +P P +T + + P+ + PT P+  
Sbjct: 3435 QTPTTTPITTTTVP-----PTPTPTGTQTPT-----TTPITTTTVPPTPTPTG--TQTP 3483

Query: 503 TYTPITSSVS-TIROYPVSAQAPNSA-ITAQTGVGVASTVHLNPMQLMTVDASHARHIQ 560  
T TPIT++ + T P Q P + IT TV T Q T  
Sbjct: 3484 TTTPTTTTVPPTPTPTGTQTPTTTPITTTTVPPTPTPT--TGTQTPTTTPITTTTVP 3541

Query: 561 IQPAPISTQGIQAPIGTPGI---QPAPLGTQGIHSATPINTQGL---QPAPMGTOQQPQ- 613  
P P TQ PI T P P GTQ + TPI T P P GTQ P  
Sbjct: 3542 PTPTPTGTQTPTTTPITTTTVPPTPTPTGTQ-TPTTTPITTTTVPPTPTPTGTQTPT 3600

Query: 614 -PEGKTSAVVLADGATIVANPISNFFSAAPAT-TVVQTHSQSASTNAPAQGSSPRPSIL 671  
P T+ V T P + P + T T T +Q+ +T ++ P+  
Sbjct: 3601 TPITTTTVPPTPTPTGTQTPTTTPITTTTVPPTPTPTGTQTPTTTPITTTTVPPTPT 3660

Query: 672 RKKPATDGAKPKEIHVSMATPVTVSMETVSNQNNDOPTIAVP---PTAQPPPTIPTMI 728  
T P + TP +T + T P PT Q P T P  
Sbjct: 3661 PTGTQTPTTTPITTTTVPPTPTPTGTQTPTTTPITTTTVPPTPTPTGTQTPTTTPITT 3720

Query: 729 AAASPPSQPAVALSTIPGAVPITPPITIAAAPPSPVTVGGLSSVLGPPVPEIKVKEEV 788  
P+ T P PIT TT P P+ T G+ + P V  
Sbjct: 3721 TTTVPPTPTPTGTQT-PTTTPIT---TTTVPPTPTPT--GTQTPTTTPITTTTVPPTPT 3774

Query: 789 EPMDIMRPVSAVPPLATNTVSPS 811  
P P+ P T TV+P+  
Sbjct: 3775 TPTGTQTPTT-PITTTTVPPT 3796

Score = 313 (47.0 bits), Expect = 4.2e-23, P = 4.2e-23  
Identities = 169/695 (24%), Positives = 245/695 (35%)

Query: 96 VVVRPYPQVQMLSTHHAVASATPVAVTAPPAHL-TPAVPLSFSEGLMKPPPKPTMPSRPI 154  
V P P T + + TV T P TP + + P P PT P  
Sbjct: 3655 VTPTPTPTGTQTPTTTPITTTTVPPTPTPTGTQTPTTTPITTTTVPPTPTPTGTQTPT 3714

Query: 155 A-PAPPSTLSLPPKVP-GQVVTMESSIPQASAI PVATISGQQGHPSNLHHIMTTNVQMS 212  
P +T P P G T T + P T +G Q P+ TT V +  
Sbjct: 3715 TTPITTTTVPPTPTPTGTQTPTTTPITTTTVPPTPTPTGTQT-PTTTPITTTTVPPT 3773

Query: 213 IIRSNAPGP---PLHIGASHLPRGAAAAA-VMSSSKVTVLRPTSQLPNAATAQPAVQHI 268  
+ P P+ + P +++ +TT T T P I  
Sbjct: 3774 PTPTGTQTPTTTPITTTTVPPTPTPTGTQTPTTTPITTTTVPPTPTPTGTQTPTTTP 3833

Query: 269 IHQIQSRPPVTTSNAI PPAVVATVSATRAQSPVITTTAAHATDSALSRPTLSIQHPPSA 328  
+ P T P T + T +P T T T + T++ P  
Sbjct: 3834 TTTTVPPTPTPTGTQTPTTTPITTTTVPPTPTPTGTQTPTTTPITTTTVP---PTPT 3890

Query: 329 AISIQRPQSRDVTTRITLPSHPALGTPKQQLHTMAQKT-IFSTGTPVAAAT--VAPILA 385  
 Q P + TT P+ GT + T + T TP T PI  
 Sbjct: 3891 PTGTQPTTTTPIITTTTIVTPTPTGTQPTTTTPIITTTTIVTPTPTGTQPTTTTPIIT 3950

Query: 386 NTNI-PSATTAGSVSHTQAPTSTIVTMT-VPSHSSHATAVTTSNIPVAKVVPQQIHTSP 443  
 T T+ P+ T G+ + T P +T T+T P+ + T TT V P T T  
 Sbjct: 3951 TTVTPTPTPTGTQPTTTTPIITTTTIVTPTPTGTQPTTTTPIITTTTIVTPTPTPTGT 4009

Query: 444 RIQPDYPAERSSLIPISGHRASPNPVMETRSDNRPSVPVQFYFL-PTYPPSAYPLAAH 502  
 + P ++ + +P P +T + + P+ + PT P+  
 Sbjct: 4010 QTPTTTTPIITTTTIVT-----PTPTPTGTQPT-----TTPIITTTTIVTPTPTPTG--TQTP 4058

Query: 503 TYTPITSSVS-TIRQYPVSAQAPNSA-ITAQTGVGVASTVHLNPMQLMTVDASHARHIQ 560  
 T TPIT++ + T P Q P + IT T V T Q T  
 Sbjct: 4059 TTTPIITTTTIVTPTPTGTQPTTTTPIITTTTIVTPTPTPT--TGTQPTTTTPIITTTTIVT 4116

Query: 561 IQPAPISTQGIQAPIGTPIGI---QPAPLGTQGIHSATPINTQGL---QPAPMGTOQPQP 614  
 P P TQ PI T P P GTQ + TPI T P P GTQ P  
 Sbjct: 4117 PTPTPTGTQPTTTTPIITTTTIVTPTPTPTGTQ-TPTTTPIITTTTIVTPTPTPTGTQPT- 4174

Query: 615 EGKTSAVVLADGATIVANPISNPFSAAPAATTVVQTHSQSASTNAPAGSSPRPSILRKK 674  
 T+ + T+ P P T ++ ++N P + S+P+ S  
 Sbjct: 4175 ---TTPITTT--TTVTPTPTPTGTQPTPTHTSTAPIAELTTSNPPESSTPQTSRSTSS 4229

Query: 675 PATDGAKPKSEIH--VSMATPVTVSMETVSNQNDQPTIAVPP-TAQQPP--PTIPTMIA 729  
 P T+ S + M+ S T + T++ PP T PP PT T  
 Sbjct: 4230 PLTESTLLSTLPPALEMTSTAPPSTPTAPTTSGGHTLSPPSTTSSPGPTRGTTTG 4289

Query: 730 AASPPSQPAVALSTI---PGAVPITPP--ITTIAAAP-PPSVTVGSSLSSVLGPPVPEI 782  
 ++S P+ V +T P P++ P I T P P SV + L+ P E+  
 Sbjct: 4290 SSSAPTPSTVQTTTSAWTPPTPLSTPSIIRTGLRYPSSVLICCVLNDYYAPGEEV 4349

Score = 279 (41.9 bits), Expect = 1.8e-19, P = 1.8e-19  
 Identities = 138/540 (25%), Positives = 194/540 (35%)

Query: 278 PVTTSNAIPPAVVATVSATRAQSPVITTTAAH-----ATDSALSRP--TLISIQHPPSAA 329  
 P+TT+ + P T + T +P+ TTT T + + P T + P  
 Sbjct: 1946 PITTITTTIVTPTPTPTGTQPTTTTPIITTTTIVTPTPTPTGTQPTTTTPIITTTTIVTPTPT 2005

Query: 330 ISIQRPAQSRDVTTRITLPSHPALGTPKQQLHTMAQKT-IFSTGTPVAAAT--VAPILAT 386  
 Q P + TT P+ GT + T + T TP T PI T  
 Sbjct: 2006 TGTQPTTTTPIITTTTIVTPTPTPTGTQPTTTTPIITTTTIVTPTPTPTGTQPTTTTPIITT 2065

Query: 387 NTI-PSATTAGSVSHTQAPTSTIVTMT-VPSHSSHATAVTTSNIPVAKVVPQQIHTSPR 444  
 T+ P+ T G+ + T P +T T+T P+ + T TT V P T T +  
 Sbjct: 2066 TTVTPTPTPTGTQPTTTTPIITTTTIVTPTPTPTGTQPTTTTPIITTTTIVTPTPTPTGTQ 2124

Query: 445 IQPDYPAERSSLIPISGHRASPNPVMETRSDNRPSVPVQFYFL-PTYPPSAYPLAAHT 503  
 P ++ + +P P +T + + P+ + PT P+ T  
 Sbjct: 2125 TPTTTPIITTTTIVT-----PTPTPTGTQPT-----TTPIITTTTIVTPTPTPTG--TQPT 2173

Query: 504 YTPITSSVS-TIRQYPVSAQAPNSA-ITAQTGVGVASTVHLNPMQLMTVDASHARHIQ 561  
 TPIT++ + T P Q P + IT T V T Q T  
 Sbjct: 2174 TTPITTTTIVTPTPTPTGTQPTTTTPIITTTTIVTPTPTPT--TGTQPTTTTPIITTTTIVT 2231

Query: 562 QPAPISTQGIQAPIGTPIGI---QPAPLGTQGIHSATPINTQGL---QPAPMGTOQPQP-- 613  
 P P TQ PI T P P GTQ + TPI T P P GTQ P  
 Sbjct: 2232 TPTPTGTQPTTTTPIITTTTIVTPTPTPTGTQ-TPTTTPIITTTTIVTPTPTPTGTQPTTT 2290

Query: 614 PEGKTSAVVLADGATIVANPISNPFSAAPAAT-TVVQTHSQSASTNAPAGSSPRPSILR 672  
 P T+ V T P + P + T T T +Q+ +T ++ P+  
 Sbjct: 2291 PITTITTTIVTPTPTGTQPTTTTPIITTTTIVTPTPTPTGTQPTTTTPIITTTTIVTPTPTPT 2350

Query: 673 KKPATDGAKPKSEIHVSMATPVTVSMETVSNQNDQPTIAVP---PTAQQPPPTIPTMIA 729  
 T P + TP +T + T P PT Q P T P  
 Sbjct: 2351 TGTQPTTTTPIITTTTIVTPTPTPTGTQPTTTTPIITTTTIVTPTPTPTGTQPTTTTPIITT 2410

Query: 730 AASPPSQPAVALSTIPGAVPITPPITTTIAAAPPPSVTVGSSLSSVLGPPVPEIKVKEEVE 789  
 P+ T P PIT TT P P+ T G+ + P V  
 Sbjct: 2411 TTVTPTPTPTGTQPTTTTPIITTTTIVTPTPTPT--GTQPTTTTPIITTTTIVTPTPTPT 2464

Query: 790 PMDIMRPVSAVPPPLATNTVSPS 811  
 P P + P T TV+P+  
 Sbjct: 2465 PTGTQPTTTTPIITTTTIVTPT 2485

Score = 265 (39.8 bits), Expect = 5.8e-18, P = 5.8e-18  
 Identities = 179/746 (23%), Positives = 257/746 (34%)

Query: 96 VVVRYPYQVQMLSTHHAVASATPVAVTAPPAHL-TPAVPLSFSEGLMKPPPKPTMPSRPI 154  
 V P P T + + T V T P TP + + P P PT P  
 Sbjct: 3678 VTPTPTPTGTQPTTTTPIITTTTIVTPTPTPTGTQPTTTTPIITTTTIVTPTPTPTGTQPT 3737

Query: 155 A-PAPPSTLSLPPKVP-GQVVTMESSIPQASAI PVATISGQQGHPSNLHHIMTTNVQMS 212  
P +T P P G T T + P T +G Q P+ TT V +  
Sbjct: 3738 TTPITTTTPTPTPTGTQTPTTPTITTTTPTPTPTGTQT-PTTPTITTTTPTPT 3796

Query: 213 IIRSNAPGP---PLHIGASHLPRGAAAAA-VMSSSKVTVLRPTSQLPNAATAQPAVQHI 268  
+ P P+ + P +++ +TT T T P I  
Sbjct: 3797 PTPGTQTPTTPTITTTTPTPTPTGTQTPTTPTITTTTPTPTPTGTQTPTTPTI 3856

Query: 269 IHQPIQSRPPVTSNAIPPAVVATVSATRAQSPVITTTAAHATDSALSRPTLSIQHPPSA 328  
+ P T P T + T +P T T T + T++ P  
Sbjct: 3857 TTTTPTPTPTGTQTPTTPTITTTTPTPTPTGTQTPTTPTITTTTPTPTPTPTPT 3913

Query: 329 AISIQRFPAQSRDVTTRITLPSHPALGTPKQQLHTMAQKT-IFSTGTPVAAAT--VAPILA 385  
Q P + TT P+ GT + T + T TP T PI  
Sbjct: 3914 PTGTQTPTTPTITTTTPTPTPTGTQTPTTPTITTTTPTPTPTGTQTPTTPTITT 3973

Query: 386 TNTI-PSATTAGSVSHSTQAPTSTIVTMT-VPSHSSHATAVTTSNIPVAKVVPQOITHTSP 443  
T T + P+ T G+ + T P +T T+ P+ + T TT V P T T  
Sbjct: 3974 TTTVTPTPTPTGTQTPTTPTITTTTPTPTPTGTQTPTTPTITTTTPTPTPTPTPT 4032

Query: 444 RIQPDYPAERSSLIPISGHRASPNPVMETRSDNRPSVPVQFYFL-PTYPPSAYPLAAH 502  
+ P ++ + +P P +T + + P+ + PT P+  
Sbjct: 4033 QTPTTPTITTTTPTPTPTPTGTQTPTPTPTITTTTPTPTPTPTPTPTPTPTPTG 4081

Query: 503 TYTPTSSVS-TIRQPVSAQAPNSA-ITAQTGVGVASTVHLNPMQLMTVDASHARHIQ 560  
T TPIT++ + T P Q P + IT T V T Q T  
Sbjct: 4082 TTTPTITTTTPTPTPTPTGTQTPTTPTITTTTPTPTPTPTPTPTPTPTPTPTPTPT 4139

Query: 561 IQPAPISTQGIQAPIGTFGI---QPAPLGTQGIHSATPINTQGL---QPAPMGTQQPQP 614  
P P TQ PI T P P GTQ + TPI T P P GTQ P  
Sbjct: 4140 PTPPTGTQTPTTPTITTTTPTPTPTGTQT-TPTTPTITTTTPTPTPTGTQTGPP 4198

Query: 615 EGKTSAVVLADGATIVANPISNPFSAAPA---ATTVVQTHSQSA-STNAPA--QGSSPRP 668  
TS +A+ T +NP P S+ P +T+ TS + ST PA S+ P  
Sbjct: 4199 T-HTSTAPIAELTT--SNP--PPESSTPQTSRSTSSPLTESTLLSTLPPAIEMTSTAPP 4253

Query: 669 SILRKKPATDGAKPKSEIHVSMATPVTVSMETVSNQNDQPTIAVPPTAQPPPTIPTMI 728  
S T G S + +P + ++ PT + T T PT  
Sbjct: 4254 STPTAPTTSGGHTLSPPTSTTSPPGTPTRGTTTGSSSAPTSTVQTTTSAWT-PTPT 4312

Query: 729 AAASPPSQPAVALSTIPGAVPITPPITIAAAPPSVTVGSSLSSVLGPPVPEIKVKEEV 788  
++P L P +V I + AP V G+ + E  
Sbjct: 4313 PLSTPSIIRTTGLRYPSSVLICCVLNDYYAPGEEV-YNGTYGDTCYFVNCLSCTLEF 4371

Query: 789 EPMIDMRPVSAPVPLATNTVSPSLALLANNLSMPTSDLPFGASPRKKPRKQOH 841  
S P + +T +PS ++ S PT P P P +Q++  
Sbjct: 4372 YNWSCTPSTPTPTPSKSTPTPSKP--SSTPSKPTPGTKPPECDFDPPRQEN 4422

Score = 254 (38.1 bits), Expect = 8.7e-17, P = 8.7e-17  
Identities = 167/697 (23%), Positives = 245/697 (35%)

Query: 115 SATPVAVTAPPAHLTPAVPLSFSEGLMKPPPK--PTMPSR-PIAPAPPSTLSLPPKV-PG 170  
S + T PP TP+ P + + PPP P+ P+ PI P P ST +LPP P  
Sbjct: 1587 SPPTITTTTTPPTTTPSPPTTTT---TPPTTTPSPPTTTPITP-PTSTTTLPPTTTPS 1642

Query: 171 QVTVTMESSIPQASAI PVATISGQQGHPSNLHHIMTTNVQMSIIRSNAPGPPLHIGASHL 230  
T + P + P T + + TT I + P PP +  
Sbjct: 1643 PPPTTTTTPPTTTPSPPTTTPSPPTTTPPTTTPSSPI--TTTPSPPTTMTTPS 1700

Query: 231 PRGAAAAVMSSSKVTVLRPTSQLPNAATAQPAVQHIIHQPIQS-RPPVTSNAIPPAV 289  
P SS +TT P+S + P P + PP TT +PP  
Sbjct: 1701 P-----TTTPSPITTTTTPSS---TTTPSPPTTMTTPSPPTTTPSPPTTMTLPPPT 1751

Query: 290 VATVSATRAQSPVITT-TAAHATDSALSRPTLSIQH---PPSAAISIQRFPAQSRDVTTR 344  
++ T P IT T + + + + P + + + S + +P ++  
Sbjct: 1752 TSSPLTTLPPIITPTTSPFSTTTPTTPCPLCNWGTGLDSGKPNFHKPGGDTLIGD 1811

Query: 345 ITLPSHPALGTPKQQLHTMAQKTIFSTGTPVAAATVAPILATN-----TIPSATTAGS 397  
+ P A + + ++ I G V ++ N I P A  
Sbjct: 1812 VCGPGWAANISCRATMYP--DVPIGQLGQTVVCDVSVGLICKNEDQKPGGVIPMAFCLNY 1869

Query: 398 VSHQAPTSTI--VTMTVPSHSSHATAVTTSNIPVAKVVPQOITHTSPRIQPDYPAERSS 455  
+ Q TMT + + + T TT+ I V T T + P ++  
Sbjct: 1870 EINVQCCCVTQPTTMTTTT-TENPTPTTTPITTTTPTPTPTGTQTPTTPTITTTT 1928

Query: 456 LIPISGHRASPNPVMETRSDNRPSVPVQFYFL-PTYPPSAYPLAAHTYPTITSSVS-T 513  
+ +P P +T + + P+ + PT P+ T TPIT++ + T  
Sbjct: 1929 TVT-----PTPTPTGTQTPT---TTPITTTTPTPTPTPTG--TQTPTTPTITTTTPT 1977

Query: 514 IRQYPVSAQAPNSA-ITAQTGVGVASTVHLNPMQLMTVDASHARHIQGIQAPAPISTQGIQ 572

Score = 243 (36.5 bits), Expect = 1.3e-15, P = 1.3e-15  
Identities = 110/406 (27%), Positives = 154/406 (37%)

Score = 189 (28.4 bits), Expect = 8.0e-09, P = 8.0e-09  
Identities = 92/374 (24%), Positives = 133/374 (35%)

773

Query: 786 EEVEPMDIMRPVSAVPLATNTVSPSL 812  
M + P + PL T + PS+  
Sbjct: 1739 PPTTMTTLPPPTTSSPLTTTLPSSI 1765

Score = 185 (27.8 bits), Expect = 1.6e-09, P = 1.6e-09  
Identities = 71/270 (26%), Positives = 99/270 (36%)

Query: 563 PAPISTQGIQAPIGTPGIQAPLGTQGIHSATP---INTQGLQAPMGTOQPQ---PEG 616  
P+P +T P P TP P T + + TP I+T P P T P P  
Sbjct: 1422 PSPPTTTTTPPTTTPS-PPITTTTTLPTTTPSPPISTT-TTPPTTTPSPPTTTPSP 1479

Query: 617 KTSAVVLADGATIVANPISNPFSAAPAATTVVQTHSQSASTNAPAQGSPPRPSILRKKPA 676  
T+ T P + P +P TT + T S +T P SP + P  
Sbjct: 1480 PTTTSPPTTTTTPPTTTP---SPFMTTPI-TTPASTTTLPTTTPSPPTTTTTPPP 1535

Query: 677 TDGAKPKSEIHVSMATPVTVMETVSNQNDQPTIAVPPTAQPPPTIPTMIAAASPPSQ 736  
T P + TP+T T + P+ P T PPPT + PS  
Sbjct: 1536 TTTTSPPT-----TTPITPPTSTTLPTTTPS-PPPTTTTTPPTTTPSPPTTTTTPSP 1588

Query: 737 PAVALSTIPGAVPITPPITTTIAAAPPVSVVGGSLSSVLGPPVPEIKVKEEVEPMDIMRP 796  
P + +T P +PP TT PPP+ T ++ + PP + P  
Sbjct: 1589 PTITTTTTPPTTTPSPPTT-TTTPPTTTPSPPTTTPITPPTSTTLPTTTPSP--PP 1645

Query: 797 VSAVPLATNTVSPSLALLANNLSMPTSDLPFGASP 832  
+ P T T SP + T+ PP +P  
Sbjct: 1646 TTTTTPPTTTPSPPTTTPSPPTTTTTPPTTTP 1681

Score = 183 (27.5 bits), Expect = 3.4e-09, P = 3.4e-09  
Identities = 91/390 (23%), Positives = 139/390 (35%)

Query: 326 PSAAISIQRPQRSQDVTR-ITLPSHPALGTPKQQLHTMAQKTIFFSTGTPVAAATVAPIL 384  
PS + P + T T PS P T T I +T TP+ T +P +  
Sbjct: 1399 PSPPTTTPSPPTTTTTLPTTTPSPPTTTTTPPTTTPSPPIITTTTTLPTTTPSPPI 1458

Query: 385 ATNTIPSATTAGSVSHTQAPTSTIVTMTVPSHSHATAVTTSNIP--VAKVVPQQITHTS 442  
+T T P TT S T P+ T + P+ ++ TT+ P + P T T  
Sbjct: 1459 STTTTTPPTTTPSPPTTTPSPPTTTPPTTTPPTTTPSPPTTTPITPPASTTTL 1517

Query: 443 PRIQPDYPAERSSLIPIGHRASP---NPVAMETRSNRP--SVPVQFYFLPTYPPSAY 497  
P P ++ P SP P+ T + P + P T PP+  
Sbjct: 1518 PTTTTPSPPTTTTTPPTTTPSPPTTTPITPPTSTTLPTTTPSPPTTTPPTTTPPTT 1577

Query: 498 PLAHAHTYTPITSSVSTIRQYPVSAQAPNSAITAQTVGVASTVHLNPMQL-MTVDASHAR 556  
P T TP ++T P + +P T T +T P +T S  
Sbjct: 1578 PSPPTTTPSPPTITTTTTPPTTTPSPPTTTPPTTTPPTTTPPTTTPPTTTPPTT 1634

Query: 557 HIQGIQAPISTQGIQAPIGTPGIQAPLGTQGIHSATPINTQGLQAPMGTOQPQPEG 616  
P+P T P P TP P P T T T P P  
Sbjct: 1635 LPPTTTPSPPTTTTTPPTTTPS--P-PTTTPSPPIITTTTTPPTTTPSPPTTTPSP 1691

Query: 617 KTSAVVLADGATIVANPISNPFSAAPAATTVVQTHSQSASTNAPAQGSPPRPSILRKKPA 676  
T+ + T ++PI+ + P+TT + +T +P SP + + P  
Sbjct: 1692 PTTTMTTPSPTTTPSSPIIT--TTTPSSTTPSPPTTMTTPSPTTTPSPPTTMTTLPP 1749

Query: 677 TDGAKPKSEIHVSMATPVTVMETVSNQNDQPTIAVPP 715  
T + P + + P +++ TS + PT P  
Sbjct: 1750 TTTSSPLT----TTLPPSITPPTTFSPFSTTTPTTPCVP 1784

Score = 176 (26.4 bits), Expect = 1.8e-07, P = 1.8e-07  
Identities = 101/402 (25%), Positives = 142/402 (35%)

Query: 345 ITLPSHPALGTPKQQLHTMAQKTIFFSTGTPVAAATVAPILATNTIPSATTAGSVSHTQAP 404  
IT PS P TP T +T +P T P T P TT + T P  
Sbjct: 1396 ITTPSPPTT-TPSPPTTTTTLPTTTPSPPTTTTTPPTTTPSPPIITTTTTLPTTTP 1454

Query: 405 TSTIVTMTVPSHSHATAVTTNIPVAKVVPQQITHSPRIQPDYPAERSSLIPIGHR 463  
+ I T T P ++ + TT+ + P P T T+P P PI+  
Sbjct: 1455 SPPISTTTTTPPTTTPSPPTTTPSPPTTTPPTTTPPTTTP--PPTTTPSPMTTTPITPP- 1511

Query: 464 ASPNPVAMETRSNRPVQFYFLPTYPPSAYPLAHTYTPITSSVSTIRQYPVSAQA 523  
AS + T PS P T PP+ P + T TPIT ST P + +  
Sbjct: 1512 ASTTTLPTTT----PSPPTTTT---TTPPTTTP-SPPTTTPITPPTSTTLPTTTPS 1563

Query: 524 PNSAITAQ----TGVGVASTVHLNPMQLMTVDASHARHIQGIQAPISTQGIQAPIGTP 579  
P T T T +T +P + T P+P +T P P TP  
Sbjct: 1564 PPPTTTTTPPTTTPSPPTTTPSPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTT 1618

Query: 580 G-----IQAPLGTQGIHSAT---PINTQGLQAPMGTOQPQPEGKTSAVVLADGATIV 630  
I P P T + T P T P P T P S +  
Sbjct: 1619 SPPTTTPITP-PTSTTTLPTTTPSPPTTTPPTTTPPTTTPSPPTTTPSPPTTTPPTTTP 1677

Query: 631 ANPISNPFSAAPAA-TTVVQTHSQSASTNAP-AQGSSPRPSILRKKPATDGAKPKSEIHV 688  
 S+P + P+ TT + T S + + +P +P + P T P  
 Sbjct: 1678 TTPSSPITTTSPPTTTMTTPSPITTTTPSSPITTTTTPSSSTTTSPSPPTTMTTPSP---T 1734

Query: 689 SMATPVTVSMETVSNQNDQPTIAVPPTAQPPPTIPTMIAAASPPSQPAVALSTIPG 746  
 + +P T +M T+ P P PPT + + P+ P V L G  
 Sbjct: 1735 TTPSPPTTMTTLPPTTTSSPLTTTLPSPITPPTSPF--STTTPTPCVPLCNWTG 1790

Score = 168 (25.2 bits), Expect = 9.3e-08, P = 9.3e-08  
 Identities = 89/387 (22%), Positives = 133/387 (34%)

Query: 448 DYPAERSSSLIPISGHRASPNPAMETRSDNRPSVPVQFYFLPTYPPSAYPLAAHTYTPI 507  
 DY + P+ +P+ T + + P P PT PS P T P  
 Sbjct: 1381 DYKIRVNCCWPMDCITTPSP---PTTTPSP--PTTTTLPPTTTPSP-PTTTTTTPPP 1434

Query: 508 TSSVS---TIRQYVSAQAPNSAITAQTGVGVASTVHLNPMQLMTVDASHARHIQGIQPA 564  
 T++ S T P+ P+ I+ T +T P T + P+  
 Sbjct: 1435 TTPSPPTTTTTLPTTTSPPISTTTTTPPTTT---PSPTTTPSPPTT-----TPS 1485

Query: 565 PISTQGIQAPIGTPIGI-QPAPLGTQGIHSATPINTQGLQPAPMGTOQFQ---PEGKTS 620  
 P +T P P TP P+ + P T P T P P T+  
 Sbjct: 1486 PPTTTTTPPTTTPSPPTTPTTPASTTTLPPTTTPSPPTTTTTPPTTTPSPPTT 1545

Query: 621 VVLADGATIVANPISNPFSAAPAAATTVVQTHSQSA-STNAPAQGS---SPRPSILRKKP 675  
 + +T P + P TT T + S +T P+ + +P P+ P  
 Sbjct: 1546 PITPPTSTTTLPTTTTPSPPTTTTTPPTTTPSPPTTTPSPPTTITTTTPPTTTPSP 1605

Query: 676 ATDGAKPKSEIHVS--MATPVTVSMETVSNQNDQPTIAVPPTAQPPPTIPTMIAAASP 733  
 T P S TP+T T + P+ P T PPPT +  
 Sbjct: 1606 TTTTTTPPTTTPSPPTTPTTPPTSTTTLPTTTPS-PPTTTTTPPTTTPSPPTTT 1664

Query: 734 PSQPAVALSTIPGAVPITPPITIAAAPPSVTVGSLSSVLGP---PVPEIKVKEEVE 789  
 PS P +T P + PITT + P ++T ++ P P  
 Sbjct: 1665 PSPPTTTTTPPTTTPSPPTTTPSPPTTMTTPSPPTTTPSSPITTTTTPSSTTTPSP 1724

Query: 790 PMDIRPVSAPPLATNTVSPSLALLANLMSPTSDLPFGASP 832  
 P + P P T +L + + T+ LPP +P  
 Sbjct: 1725 PTTMTTPSPPTTTPSPPTTMTTLPPTTSSPLTTTLPSPITP 1767

Score = 154 (23.1 bits), Expect = 2.7e-06, P = 2.7e-06  
 Identities = 70/277 (25%), Positives = 92/277 (33%)

Query: 565 PISTQGIQAPIGTPIGIQAPLGTQGIHSATPINTQGLQPAPMGTOQFQPEGKTS 624  
 PIST P P TP P P T + TP P T P P T +  
 Sbjct: 1457 PISTT-TTPPTTTPS--P-PTTTPSPPTTTPSPPTTTTTPPTTTPSPPTTTP--ITP 1510

Query: 625 DGATIVANPISNPFSAAPAAATTVVQTHSQSASTNAP---AQGSSPRPSILRKKPATDGA 680  
 +T P + P TT T + S T P ++ P+ P T  
 Sbjct: 1511 PASTTTLPTTTPSPPTTTTTPPTTTPSPPTTTPITPPTSTTTLPTTTPSPPTTT 1570

Query: 681 KPKSEIHVSMATPVTVSMETVSNQNDQPTIAVPPTAQ--PPPTIPTMIAAASPPSQPA 738  
 P S T T S T++ T PPT PPPT T + P P  
 Sbjct: 1571 TTPPTTTPSPPTTTPSPPTTITTTTPPTTTPSPPTTTTTPPTT-TPSPPTTTPITP 1629

Query: 739 VALSTIPGAVPITPPITIAAAPPSVTVGSLSSVLGPVPEIKVKEEVEPMDIRPVS 798  
 + +T+P +PP TT PPP+ T ++ PP+ +  
 Sbjct: 1630 TSTTTLPPTTTPSPPTT-TTTPPTTTPSPPTTTPSPPTTITTTTPPTTTPSSPITTT 1688

Query: 799 AVPPLATNTV----SPSLALLANL--SMPTSDLPFGASPRKKP 836  
 PP T T +PS + S T PP P  
 Sbjct: 1689 PSPPTTMTTTPSPPTTTPSPITTTTTPSSTTTPSPPTTMTTTPSP 1733

Score = 148 (22.2 bits), Expect = 1.1e-05, P = 1.1e-05  
 Identities = 62/254 (24%), Positives = 89/254 (35%)

Query: 583 PAPLGTQGIHSATPINTQGLQPAPMGTOQFQPEGKTS 637  
 P+P T S P T L P T P P T+ + T P+  
 Sbjct: 1399 PSPPTTTP--SPPTTTTTLPP---TTTPSPPTTTTTPPTTTPSPPTTITTTTLPPT 1452

Query: 638 FSAAPAAATTVVQTHSQSASTNAPAQGSSPRPSILRKKPATDGAKPKSEIHVS--MATPVT 695  
 + P +TT T + + + P SP P+ P T P S M TP+T  
 Sbjct: 1453 TSPPISTTT--TTPPTTTPSPPTTTPSP-PTTTPSPPTTTTTPPTTTPSPPTTTPIT 1509

Query: 696 VSMETVSNQNDQPTIAVPPTAQPPPTIPTMIAAASPPSQPAVALSTIPGAVPITPPIT 755  
 T + P+ T PP T P+ + P P + +T+P +PP T  
 Sbjct: 1510 PPASTTLPPTTTPSPPTTTTTPPTTTPS--PPTTTPITPPTSTTTLPTTTPSPPT 1567

Query: 756 TIAAAPPSVTVGSLSSVLGPVPEIKVKEEVEPMDIRPVSAPPLATNTVSPSLALL 815  
 T PPP+ T ++ PP + PP T P+ +  
 Sbjct: 1568 T-TTTPPTTTPSPPTTTPSPPTTITTTTPPTTTPSPPTTTTTPPTTTPSPPTTTPIT 1626

Query: 816 ANNLSMPTSDLPPGASPRKKP 836  
 S T+ LPP +P P  
 Sbjct: 1627 TPPTS--TTTLPTTTTPSPPP 1645

Score = 131 (19.7 bits), Expect = 1.2e-03, P = 1.2e-03  
 Identities = 112/492 (22%), Positives = 174/492 (35%)

Query: 96 VVVRPYQVQMLSTHHAVASATPVAVTAPPAHL-TPAVPLSFSEGLMKPPPKPTMPSRPI 154  
 V P P T + + T V T P TP + + P P PT P  
 Sbjct: 3977 VTPTPTGTGTPTTTPIITTTTPTPTPTGTGTPTTTPIITTTTPTPTPTGTGTPT 4036

Query: 155 A-PAPPSTLSLPPKVP-GQVTVMESIPQASAI PVATISGOQGHPSNLHHIMTTNVQMS 212  
 P +T P P G T T + P T +G Q P+ TT V +  
 Sbjct: 4037 TTPITTTTPTPTPTGTGTPTTTPIITTTTPTPTPTGTGT-PTTPIITTTTPTPT 4095

Query: 213 IIRSNAPGP---PLHIGASHLPRGAAAA-VMSKKVTVLRPTSQLPNAATAQPAVQHI 268  
 + P P+ + P +++ +TT T T P I  
 Sbjct: 4096 PTPTGTGTPTTTPIITTTTPTPTPTGTGTPTTTPIITTTTPTPTPTGTGTPTTTPI 4155

Query: 269 IHQPIQSRPVTTSNAIPPA--VVATVSATRAQSPVITTA--AHATDSALSRLTSLIQH 324  
 + P T P + T + T +P T T H + + + T S  
 Sbjct: 4156 TTTTPTPTPTGTGTPTTTPIITTTTPTPTPTGTGTGPPTHSTAPIAELTTSNPP 4215

Query: 325 PPSAAISIQRPAQS--RDVTTRI-TLPSHPALGTPKQQLHTMAQKTI FSTGTEVAAATVA 381  
 P S+ R S + TT + TLP PA+ + T T + T T++  
 Sbjct: 4216 PESSTPQTSRSTSSPLTESTTLLSTLP--PAI----EMTSTAPPSTPTAPTTTSGGHTLS 4269

Query: 382 PILANTNIPSAT-TAGSVS-HTQAPTSTIVMTVPSSHSHATAVTTSNIPVAKVVPQKIT 439  
 P +T T P T T G+ + + APT + V T S A T + P++ P I  
 Sbjct: 4270 PPPSTTSPFGTPTGTGTGSSSAPTSTVQTTTS-----AWTPTPTPLS--TPSIIR 4321

Query: 440 HTSPRIQPDYPAERSSLIPISGHRASPNP-VAMETRSDN----RPSVPVQFQYFLPTYP- 493  
 T ++P YP+ ++ +P V T D S+ +++ + P  
 Sbjct: 4322 TTG--LRP-YPSSVLICCVLNDYYAPGEEVYNGTYGOTCYFVNCSLSCTLEFYNWSCPS 4378

Query: 494 -PSAYPLAAHTYTPITSSVSTIRQYVSAQAPNSAITAQTGVGVAHVHLPMLMTVDA 552  
 PS P + + TP S S+ P P T L + T  
 Sbjct: 4379 TPSPTPTPSKS-TPTSPKPSSTPSKPTPGTKPECPDFDPPRQENETWWLDCDFMATCKY 4437

Query: 553 SHARHIQIQI---PAPISTQGIQAPIGTP 579  
 ++ I ++ P P + G+QP + P  
 Sbjct: 4438 NNTVEIVKVECEPPMPPTCSNGLQPVVRVEDP 4468

Score = 117 (17.6 bits), Expect = 1.8e-02, P = 1.8e-02  
 Identities = 41/156 (26%), Positives = 55/156 (35%)

Query: 710 TIAVPPTAQPPPTIPTMIAAASPSPQPAVALSTIPGAVPITPPITIAAAPPSVTVGG 769  
 T + P T PPPT T + + PS P +T P +PITT P P+ T  
 Sbjct: 1398 TPSPPTTTPSPPTTTTLPTTTPSPPTTTTTPPTTTPSPPTT-TTTLPTTTPSP 1456

Query: 770 SLSSVLGPPVPEIKVKEEVEPMDIMRPVSAVPLATNTVPSLALLANNLSMPTSDLPPG 829  
 +S+ PP P P + P T T SP T+ PP  
 Sbjct: 1457 PISTTTTP-----PTTTPSPPTTTPSPPTTTPSPPTTTTTP-PTTTPSPPM 1504

Query: 830 ASPRKKPRKQHVISTEEDMMETNSDDEKSTAKS 865  
 +P P + T T+T +T S  
 Sbjct: 1505 TTPITPPASTTTLPTTTPSPPTTTTTPPTTTPS 1540

Score = 61 (9.2 bits), Expect = 1.6e-09, P = 1.6e-09  
 Identities = 23/93 (24%), Positives = 41/93 (44%)

Query: 397 SVSHTQAPTSTIVMTVPSSHSHATAVTTSNIPVAKVV---PQIHTSPRIQPDYPAE 452  
 S++ + +T T+T+P+ + T TT+ P + V P+ S I D+P+  
 Sbjct: 1257 SITTRPSTLTFTTTITLPTTPTSTTTTTTTTPTSSSTVLSTTPKLCCLWSDWINEDHPSS 1316

Query: 453 RSS---LIPISGHRASPNPVAMETRSDNRPSVPVQ 484  
 S P G +P + E RS P + +  
 Sbjct: 1317 GSDDGDREPFQVCGAPEDI--ECRSVKDPHLSLE 1349

Score = 50 (7.5 bits), Expect = 8.0e-09, P = 8.0e-09  
 Identities = 16/41 (39%), Positives = 19/41 (46%)

Query: 334 RPAQSRDVTTRITLPSHPALGTPKQQLHTMAQKTI FSTGTP 374  
 RP+ TT ITLP+ P T T T+ ST TP  
 Sbjct: 1261 RPSTLTFTTT-ITLPTTPTSTTTTTTTTPTSSSTVLST-TP 1299

Score = 46 (6.9 bits), Expect = 5.4e-08, P = 5.4e-08  
 Identities = 24/106 (22%), Positives = 37/106 (34%)

Query: 324 HPPSAAISIQRPAQSRDVTTRITLPSHPALGTPKQQLHTMAQKTI FSTGTPVAAATVAPI 383  
 +PP A++ + +S T + P G Q A G I



Sbjct: 1196 YPPGASVPTEETCKSCVCTNSSQVVCREEGKILNQTQDGAFCYWEICGPNGTVEKHFNI 1255

Query: 384 LATNTIPSA-TTAGSVSHTQAPTSTIVTMTVPSSHSHATAVTTSNI 428  
+ T P S T T + + + P T S T T + + S T T +

Sbjct: 1256 CSITTRPSTLTFTTITLPTTPTSTTTTTTTTTPTSTVLSTTPKL 1301

Score = 44 (6.6 bits), Expect = 8.7e-08, P = 8.7e-08  
Identities = 14/34 (41%), Positives = 17/34 (50%)

Query: 478 RPSVPVQFQYF-LPTYPPSAYPLAAHTYTPITSSV 511  
R P S F L P T P S + T T P + S + V

Sbjct: 1261 RPSTLTFTTITLPTTPTSTTTTTTTTTPTSTV 1294

Pedant information for DKFZphtes3\_2a11, frame 2

Report for DKFZphtes3\_2a11.2

[LENGTH] 1048  
[MW] 110324.04  
[pI] 9.83  
[HOMOL] PIR:I47141 gastric mucin (clone PGM-2A) - pig (fragment) 8e-15  
[FUNCAT] 30.90 extracellular/secretion proteins [S. cerevisiae, YIR019c] 1e-09  
[FUNCAT] 30.01 organization of cell wall [S. cerevisiae, YIR019c] 1e-09  
[FUNCAT] 01.05.01 carbohydrate utilization [S. cerevisiae, YIR019c] 1e-09  
[FUNCAT] 30.02 organization of plasma membrane [S. cerevisiae, YDR420w] 4e-09  
[FUNCAT] 01.05.04 regulation of carbohydrate utilization [S. cerevisiae, YDR420w] 4e-09  
[FUNCAT] 98 classification not yet clear-cut [S. cerevisiae, YJR151c] 4e-06  
[FUNCAT] 03.04 budding, cell polarity and filament formation [S. cerevisiae, YGR014w] 1e-05  
[FUNCAT] 11.01 stress response [S. cerevisiae, YHL028w] 1e-04  
[FUNCAT] 09.01 biogenesis of cell wall [S. cerevisiae, YHL028w] 1e-04  
[EC] 3.2.1.3 Glucan 1,4-alpha-glucosidase 3e-08  
[PIRKW] glycosidase 3e-08  
[PIRKW] transmembrane protein 3e-08  
[PIRKW] polysaccharide degradation 3e-08  
[PIRKW] glycoprotein 9e-08  
[PIRKW] calcium binding 9e-08  
[PIRKW] hydrolase 3e-08  
[PIRKW] cytoskeleton 7e-08  
[SUPFAM] equine herpesvirus glycoprotein X 2e-07  
[SUPFAM] yeast glucan 1,4-alpha-glucosidase homolog 3e-08  
[SUPFAM] polymorphic epithelial mucin 7e-08  
[SUPFAM] glucan 1,4-alpha-glucosidase homology 3e-08  
[SUPFAM] equine herpesvirus 1 glycoprotein homology 2e-07  
[PROSITE] MYRISTYL 9  
[PROSITE] AMIDATION 1  
[PROSITE] CAMP\_PHOSPHO\_SITE 2  
[PROSITE] CK2\_PHOSPHO\_SITE 10  
[PROSITE] PKC\_PHOSPHO\_SITE 12  
[PROSITE] ASN\_GLYCOSYLATION 3  
[KW] Irregular  
[KW] LOW\_COMPLEXITY 20.04 %

SEQ MGPPRHPQAGEIEAGGAGGRRRLQVEMSSQFPRLGAPSTGLSQAPSQIANSGSAGLINP  
SEG .....xxxxxxxxxxxxx.....  
PRD ccc

SEQ AATVNDSEGRDSEVSAREHMSSSSSLQSRREEKQEPVVVRPYPQVQMLSTHHAVASATPVA  
SEG .....xxxxx.....xxxxxxxxxxxxx  
PRD ccc

SEQ VTAPPAHLTPAVPLSFSEGLMKPPPKPTMPSRPIAPAPPSTLSLPPKVPQGVTVTMESSI  
SEG xxxxxxxxxxxxxxxx.....xxxxxxxxxxxxx.....  
PRD ccc

SEQ PQASAI PVATISGQQGHPSNLHHIMTTNVQMSIIRSNA GPPLHIGASHLPRGAAAAAVM  
SEG .....xxxxx..  
PRD ccc

SEQ SSSKVTTVLRPTSQLPNAATAQPAVQHIIHQPIQSRPPVTTNSAIPPAVVATVSATRAQS  
SEG .....  
PRD ccc

SEQ PVITTTAAHATDSALSRPTLSIQHPPSAAISIQRPASQSRDVTTRITLPSHPALGTPKQQL  
SEG .....  
PRD ccc

```

SEQ HTMAQKTI FSTGTPVAAATVAPILATNTIPSATTAGSVSHTQAPTSTIVTMTVPSHSSHA
SEG .....XXXXXXXXXX.....
PRD cccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc

SEQ TAVTTSNIPVAKVVPQQITHTSPIRQPDYPAERSSLIPISGHRASPNPVAMETRSDNRPS
SEG .....XXXXXXXXXX.....
PRD cccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc

SEQ VPVQFYFLPTYPSAYPLAAHTYTPITSSVSTIRQYPVSAQAPNSAITAQTVGVGVASTV
SEG .....
PRD cccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc

SEQ HLNPMQLMTVDASHARHIQGIQPAPISTQGIQPAPIGTGPIQAPPLGTQGIHSATPINTQ
SEG .....
PRD cccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc

SEQ GLQPAPMGTTQQPQPEGKTSAVVLADGATIVANPISNPFSAAPAATTVVQTHSQSASTNAP
SEG .....
PRD cccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc

SEQ AQGSSPRPSILRKKPATDGAKPKSEIHVSMATPVTVMETVSNQNNQPTIAVPPTAQQP
SEG .....XXXXXXXXXX.....
PRD cccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc

SEQ PPTIPTMIAAASPPSQPAVALSTIPGAVPITPPITTIAAAPPSVTVGGLSSVLGPPVP
SEG .....XXXXXXXXXX.....
PRD cccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc

SEQ EIKVKEEVEPMDIMRPVSAVPLATNTVSPSLALLANNLSMPTSDLPGASPRKKPRKQO
SEG .....XXXXXXXXXX.....
PRD cccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc

SEQ HVISTEEGDMETNSTDDEKSTAKSLLVKAERKSPKEYIDEEGVRYVPVRPRPPITLL
SEG .....XXXXXXXXXX.....
PRD cccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc

SEQ RHYRNPWKAAYHHFQRYSDVRVKEEKAMLQEIANKGVSCRAQGWKVHLCAAQLQLTN
SEG .....
PRD eccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc

SEQ LEHDVYERLTNLQEGIIPKKAATDDDLHRINELIQNMQRCKLVMDQISEARDSMLKVL
SEG .....
PRD cccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc

SEQ DHKDRVLKLLNKGTVKKVSKLRKEKV
SEG .....XXXXXXXXXX.....
PRD hhhhhhhhhccccccccccccccccccccccccccccccccccccccccccccccc

```

## Prosites for DKFzphes3\_2all.2

PS00001	818->822	ASN_GLYCOSYLATION	PDOC00001
PS00001	854->858	ASN_GLYCOSYLATION	PDOC00001
PS00001	1033->1037	ASN_GLYCOSYLATION	PDOC00001
PS00004	872->876	CAMP_PHOSPHO_SITE	PDOC00004
PS00004	1037->1041	CAMP_PHOSPHO_SITE	PDOC00004
PS00005	68->71	PKC_PHOSPHO_SITE	PDOC00005
PS00005	75->78	PKC_PHOSPHO_SITE	PDOC00005
PS00005	242->245	PKC_PHOSPHO_SITE	PDOC00005
PS00005	342->345	PKC_PHOSPHO_SITE	PDOC00005
PS00005	355->358	PKC_PHOSPHO_SITE	PDOC00005
PS00005	442->445	PKC_PHOSPHO_SITE	PDOC00005
PS00005	513->516	PKC_PHOSPHO_SITE	PDOC00005
PS00005	665->668	PKC_PHOSPHO_SITE	PDOC00005
PS00005	831->834	PKC_PHOSPHO_SITE	PDOC00005
PS00005	862->865	PKC_PHOSPHO_SITE	PDOC00005
PS00005	940->943	PKC_PHOSPHO_SITE	PDOC00005
PS00005	1035->1038	PKC_PHOSPHO_SITE	PDOC00005
PS00006	63->67	CK2_PHOSPHO_SITE	PDOC00006
PS00006	68->72	CK2_PHOSPHO_SITE	PDOC00006
PS00006	75->79	CK2_PHOSPHO_SITE	PDOC00006
PS00006	88->92	CK2_PHOSPHO_SITE	PDOC00006
PS00006	135->139	CK2_PHOSPHO_SITE	PDOC00006
PS00006	473->477	CK2_PHOSPHO_SITE	PDOC00006
PS00006	844->848	CK2_PHOSPHO_SITE	PDOC00006
PS00006	855->859	CK2_PHOSPHO_SITE	PDOC00006
PS00006	959->963	CK2_PHOSPHO_SITE	PDOC00006
PS00006	984->988	CK2_PHOSPHO_SITE	PDOC00006
PS00008	15->21	MYRISTYL	PDOC00008

WO 01/12659

PCT/IB00/01496

PS00008	16->22	MYRISTYL	PDOC00008
PS00008	36->42	MYRISTYL	PDOC00008
PS00008	233->239	MYRISTYL	PDOC00008
PS00008	372->378	MYRISTYL	PDOC00008
PS00008	533->539	MYRISTYL	PDOC00008
PS00008	535->541	MYRISTYL	PDOC00008
PS00008	590->596	MYRISTYL	PDOC00008
PS00008	768->774	MYRISTYL	PDOC00008
PS00009	19->23	AMIDATION	PDOC00009

(No Pfam data available for DKF2phtes3\_2a11.2)

DKFZphtes3\_2a17  
-----

group: metabolism

DKFZphtes3\_2a17 encodes a novel 574 amino acid protein without similarity to known proteins.

The novel protein contains a thiol protease cys pattern. Eukaryotic thiol proteases (EC 3.4.22.-) are a family of proteolytic enzymes containing an active site cysteine. Cathepsins belong to this protease family.

The new protein can find application in modulation of proteolytic processes and as a new enzyme for proteomic analysis and biotechnologic production processes.

unknown

complete cDNA, complete cds, EST hits

Sequenced by EMBL

Locus: unknown

Insert length: 2312 bp

Poly A stretch at pos. 2300, polyadenylation signal at pos. 2273

```

1  GTTTTCACCT GATCATTAGA AACTAATGAA ACACCTTTTA AGTCTTATGA
51  ATTCAGGTGA CACTGTTTTC CAGATGCCCTT GGCAGCTGGT ACAGGGCCTC
101 TGA AAAAATGG AACCAAATTC TCTGAGGACT AAAGTCCCAG CTTTCTTATC
151 TGATTTGGGG AAGCCACAT  TGAGGGGAAT CAGAAAGTGT CCCCGATGTG
201 GCACATACAA TGGAAACCCGG GGACTGAGCT GTAAGAACAA GACATGTGGA
251 ACCATATTCC GCTACGGTGC ACGCAAGCAG CCTAGTGTG AAGCTGTCAA
301 AATCATTACA GGCTCTGATC TTCAGGTCTA CTCAGTGGCG CAAAGAGACC
351 GGGGCCCTGA TTACCGATGC TTTGTGGAGC TCGGGGTTTC AGAGACAACA
401 ATCCAGACAG TGGATGGGAC GATCATCACT CAGCTGAGCT CTGGACGGTG
451 TTATGTCCCC TCATGCCTGA AAGCTGCCAC TCAAGGCGTT GTGGAAAACC
501 AGTGCCAGCA CATCAAGCTG GCGGTGAACT GCCAGGCAGA GGCCACCCCT
551 CTGACCCCTG AGAGCTCGGT CCTGAATGCA ATGCAGGCCT CCCCGAAAC
601 CAAACAGACC ATCTGGCAGT TGGCCACGGA ACCCACAGGT CCTCTGGTGC
651 AGAGAATTAC TAAAAACATC TTGGTGGTGA AATGCAAGGC AAGCCAGAA
701 CACAGTTTGG GGTATTTGCA TACATCTTTT GTGCAGAAAG TCAGTGGCAA
751 AAGCTTGCTC GAGCGCCGCT TCTTCTGCTC CTGTCAAGCT CTGAAATCGC
801 ACAAGTCAAA TGCTTCCAAG GATGAGACAG CCCAGAGATG CATTCAATTC
851 TTTGCTTGCA TCTGTGCCTT TGCCAGTGAT GAGACACTGG CTCAGGAATT
901 CTCAGACTTC CTAAATTTTG ATTCCAGCGG TCTTAAAGAG ATTATGTAC
951 CCCAGTTAGG TTGCCATTCA GAATCAACAG TATCTGCTTG TGAGTCTACT
1001 GCCTCTAAGT CAAAGAAGAG GAGAAAGGAT GAAGTATCTG GTGCACAGAT
1051 GAACAGTTCA CTACTGCCTC AAGATGCAGT GAGCAGTAAT CTAAGGAAAA
1101 GTGGCCTGAA AAGCCTGTG GTTGCTTCCT CGTTAAAAAG GCAGGCCTGT
1151 GGTGAGCTGT TAGATGAGGC ACAAGTGACT TTATCCTTCC AAGACTGGCT
1201 GGCCAGTGTC ACAGAACGCA TCCATCAAAC CATGCACTAT CAGTTTGATG
1251 GCAAACCCAGA ACCATTGGTG TTCCACATTG CTCAGTCATT TTTGATGCC
1301 CTGCAACAAA GAATATCTAT AGGAAGTGCA AAAAAACGGC TCCCAACTC
1351 CACCACAGCT TTTGTTCCGA AAGATGCCCT GCCACTGGGA ACCTTTTCCA
1401 AGTATACTTG GCATATCACT AATATCCTGC AAGTTAAACA AATCTTAGAT
1451 ACCCCAGAGA TGCCCTTGGA AATCACCCGT AGCTTTATCC AGAACCAGAG
1501 TGGGACTTAT GAGCTATTTA AATGCCCTAA AGTGGAAGTA GAAAGCATAG
1551 CAGAAACCTA CGTCTGTATA GAAAAACAAC CAGTGCTGCG ACCCTTGGA
1601 CTAAAAACTT TTCTCAAAGT TGGCAACACT TCCCAGATC AAAAGGAGCC
1651 AACACCTTTC ATCATCGAGT GGTATCCAGA TATCCTTCCC CAATCTAAGA
1701 TTGGCGAGCT GCGGATCAAG TTTGAGTATG GCCACCACCG GAATGGGCAT
1751 GTGGCGGAGT ACCAAGACCA GCGGCCCCCC TTGGACCAGC CCTTGGAACT
1801 GGCCCTCTG ACCACTATTA CTTTCCCTTA AAGCAAAACA AGATAATAT
1851 CTTTTGCTGC TTAATTTGCA CATCCCCACC CCTTGACAAC TTTAATGCT
1901 AGTTAGGCAC TTAGATGGCC CTGTTCCTTG GTAACTGCT CTTAGCTAAG
1951 ATGCAAAATC TCAGTGCTTT CAAGTGGATT CTGTTGAAGA AAATCTCTTG
2001 TAAATAGCCT TTTGATGCT GCTGTGTACA GTCTTCATTA TGCATTGGGC
2051 AGTATTTCTG GCTAGAGTTT TAAAGGAAC AGAAAGAAAA CCAGCTTATT
2101 TTCCTTCTTA CGGACTCATC TTAGCGTTT ATTTCAACCT TTTGCTAATT
2151 CTCTGAGAAA TCTGCAGCAC TCAGCCATAC ACCAACAGTG TTGGAAAAGT
2201 AACACCTTGG TTAGGGCAGA ATGTTAAAGA CCATCTTGGC AGAGTTCCAG
2251 CCACGCTCTT TATTCTGTTT TCAAATAAAG CAGTGTCACT AGTTTTCTCT
2301 AAAAAAAAAA AA
```

#### BLAST Results

No BLAST result

## Medline entries

No Medline entry

## Peptide information for frame 2

ORF from 107 bp to 1828 bp; peptide length: 574  
 Category: putative protein

```

1  MEPNSLR TKV  PAFLSDLGKA  TLRGIRKCP  CGTYNGTRGL  SCKNKTCGTI
51  FRYGARKQPS  VEA VKIITGS  DLQVYSVRQR  DRGPDYRCFV  ELGVSETTIQ
101 TVDGTIITQL  SSGRCYVPSC  LKAATQGVVE  NQCQHIKLV  NCQAEATPLT
151 LKSSVLNAMQ  ASPETKQTIW  QLATEPTGPL  VQRITKNILV  VKCKASQKHS
201 LGYLHTSFVQ  KVSQKSLPER  RFFCSCQTLK  SHKSNASKDE  TAQRCHFFA
251 CICAFAFDET  LAQEFSDFLN  FDSSGLKEII  VPQLGCHSES  TVSACESTAS
301 KSKKRRKDEV  SGAQMNSSL  PQDAVSSNLR  KSGLKKPVVA  SSLKRQACGQ
351 LLDEAQVTL  FQDWLASVTE  RIHQTMHYQF  DGKPEPLVFH  IQSFFDALQ
401 QRISIGSAKK  RLPNSTAFV  RKDALPLGTF  SKYTWHITNI  LQVKQILDTP
451 EMPLAITRSF  IQNRDGTIEL  FKCPKVEVES  IAETYGRIEK  QPVLRLPELK
501 TFLKVGNTSP  DQKEPTPFII  EWIPDILPQS  KIGELRIKFE  YGHRHRGHVA
551 EYQDQRPPLD  QPLELAPLTT  ITFP

```

## BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphtes3\_2a17, frame 2

No Alert BLASTP hits found

Pedant information for DKFZphtes3\_2a17, frame 2

## Report for DKFZphtes3\_2a17.2

```

[LENGTH]      574
[MW]           64076.89
[pI]           9.15
[PROSITE]      MYRISTYL           5
[PROSITE]      CK2_PHOSPHO_SITE      9
[PROSITE]      PKC_PHOSPHO_SITE     14
[PROSITE]      ASN_GLYCOSYLATION     5
[PROSITE]      THIOL_PROTEASE_CYS    1
[KW]           Alpha_Beta

```

```

SEQ  MEPNSLR TKVPAFLSDLGKATLRGIRKCPRCGTYNGTRGLSCKNKTCGTIFRYGARKQPS
PRD  cccccccccchhhhhccccchhhhhcccccccccccccccccccccccccccccccccccc

SEQ  VEA VKIITGS DLQVYSVRQRDRGPDYRCFVELGVSETTIQTVDGTIITQLSSGRCYVPSC
PRD  cccccccccccccccccccccccccccccccccccccccccccccccccccccccccchh

SEQ  LKAATQGVVENQCQHIKLVNCQAEATPLTLKSSVLNAMQASPETKQTIWQLATEPTGPL
PRD  hhhhhhhhhccccchhhhhheehhhhhhhccccchhhhhhhhhccccchhhhhhhhhccccch

SEQ  VQRITKNILVVKCKASQKHS LGYLHTSFVQKVSQKSLPERFFCSCQTLKSHKSNASKDE
PRD  hhhhhhhheeeeecccccccccccccccccccccccccccccccccccccccccccccccc

SEQ  TAQRCHFFACICAFAFDET LAQEFSDFLNFDSSGLKEIIVPQLGCHSESTVSACESTAS
PRD  hhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh

SEQ  KSKKRRKDEVSGAQMNSSLLPQDAVSSNLRKSGLKKPVVASSLKRQACGQLLDEAQVTL
PRD  ccchhhhhccccccccccccccccccccchhhhhhhccccceehhhhhhhhhchhhhhhhhh

SEQ  FQDWLASVTERI HQTMHYQFDGKPEPLVFHI PQSFFDALQQRISIGSAKKRLPNSTAFV
PRD  hhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh

SEQ  RKDALPLGTFSKYTWHITNILQVKQILDTPEMPLAITRSFIQNRDGTIELFKCPKVEVES
PRD  eccccccccccccccccccccccccccccccccccccccccccccccccccccccccceehh

SEQ  IAETYGRIEKQPVLRLPELKTFLKVGNTSPDQKEPTPFII EWIPDILPQSKIGELRIKFE
PRD  hhhhhhhhhcccccccccccccccccccccccccccccccccccccccccccccccccccc

```

SEQ YGHRNGHVAEYQDQRPPLDQPLELAPLTITFP  
 PRD ecccccccccccccccccccccccccccccccccc

## Prosites for DKFZphtes3\_2a17.2

PS00001	35->39	ASN_GLYCOSYLATION	PDOC00001
PS00001	44->48	ASN_GLYCOSYLATION	PDOC00001
PS00001	235->239	ASN_GLYCOSYLATION	PDOC00001
PS00001	316->320	ASN_GLYCOSYLATION	PDOC00001
PS00001	414->418	ASN_GLYCOSYLATION	PDOC00001
PS00005	5->8	PKC_PHOSPHO_SITE	PDOC00005
PS00005	21->24	PKC_PHOSPHO_SITE	PDOC00005
PS00005	41->44	PKC_PHOSPHO_SITE	PDOC00005
PS00005	76->79	PKC_PHOSPHO_SITE	PDOC00005
PS00005	112->115	PKC_PHOSPHO_SITE	PDOC00005
PS00005	150->153	PKC_PHOSPHO_SITE	PDOC00005
PS00005	196->199	PKC_PHOSPHO_SITE	PDOC00005
PS00005	213->216	PKC_PHOSPHO_SITE	PDOC00005
PS00005	228->231	PKC_PHOSPHO_SITE	PDOC00005
PS00005	231->234	PKC_PHOSPHO_SITE	PDOC00005
PS00005	302->305	PKC_PHOSPHO_SITE	PDOC00005
PS00005	342->345	PKC_PHOSPHO_SITE	PDOC00005
PS00005	369->372	PKC_PHOSPHO_SITE	PDOC00005
PS00005	407->410	PKC_PHOSPHO_SITE	PDOC00005
PS00006	68->72	CK2_PHOSPHO_SITE	PDOC00006
PS00006	216->220	CK2_PHOSPHO_SITE	PDOC00006
PS00006	237->241	CK2_PHOSPHO_SITE	PDOC00006
PS00006	293->297	CK2_PHOSPHO_SITE	PDOC00006
PS00006	360->364	CK2_PHOSPHO_SITE	PDOC00006
PS00006	367->371	CK2_PHOSPHO_SITE	PDOC00006
PS00006	394->398	CK2_PHOSPHO_SITE	PDOC00006
PS00006	480->484	CK2_PHOSPHO_SITE	PDOC00006
PS00006	508->512	CK2_PHOSPHO_SITE	PDOC00006
PS00008	32->38	MYRISTYL	PDOC00008
PS00008	93->99	MYRISTYL	PDOC00008
PS00008	104->110	MYRISTYL	PDOC00008
PS00008	127->133	MYRISTYL	PDOC00008
PS00008	312->318	MYRISTYL	PDOC00008
PS00139	109->121	THIOL_PROTEASE_CYS	PDOC00126

(No Pfam data available for DKFZphtes3\_2a17.2)

DKF2phtes3\_2d15

group: testes derived

DKF2phtes3\_2d15 encodes a novel 274 amino acid protein with similarity to *C.elegans* Cosmid F25H2.1.

The novel protein contains a Pfam predicted C2-domain.  
No informative BLAST results; No predictive prosite, pfam or SCOP motive.

The new protein can find application in studying the expression profile of testis-specific genes.

similarity to *C.elegans* F25H2.1

complete cDNA, complete cds, EST hits

Sequenced by EMBL

Locus: unknown

Insert length: 3615 bp

Poly A stretch at pos. 3603, polyadenylation signal at pos. 3578

```
1 GCGGGCGGCT CGAGGTGACA ACTGTCTCCG TCGCAGGCTC CGGCGGGGGC
51 GCAGGAGGTC GCCCGGCGCG TCACTGTCCG GTCCGGCAGC CACGGGGGCC
101 GCCGAGCAC CATGGCGACC ACCGTCAGCA CTCAGCGCGG GCCGGTGTAC
151 ATCGGTGAGC TCCCGCAGGA CTTCCTCCGC ATCAGGCCCA CACAGCAGCA
201 GCGCAGGTC CAGCTGGACG CCCAGCGCGC CCAGCAGCTG CAGTACGGAG
251 GCGCAGTGGG CACCGTGGGC CGACTGAACA TCACGGTGGT ACAGGCAAAAG
301 TTGGCCAAGA ATTACGGCAT GACCCGCGAT GACCCCTACT GCCGACTGCG
351 CCTGGGCTAC GCGGTGTACG AGACGCCAC GGCACACAAT GGCGCCAAGA
401 ATCCCGCTG GAATAAGGTC ATCCACTGCA CGGTGCCCCC AGGCGTGGAC
451 TCTTTCTATC TCGAGATCTT CGATGAGAGA GCCTTCTCCA TGGACGACCG
501 CATTGCCTGG ACCCATCA CAATCCCGGA GTCCCTGAGG CAGGGCAAGG
551 TGGAGGACAA GTGGTACAGC CTGAGCGGGA GGCAGGGGGA CGACAAGGAG
601 GGCATGATCA ACCTCGTCAT GTCTACGCG CTGCTTCCAG CTGCCATGGT
651 GATGCCACCC CAGCCCGTGG TCCTGATGCC AACAGTGTAC CAGCAGGGCG
701 TTGGCTATGT GCCCATCACA GGGATGCCCG CTGTCTGTAG CCCCGGCATG
751 GTGCGCGTGG CCCTGCCCCC GGC CGCGGTG AACGCCAGC CCCGCTGTAG
801 CGAGGAGGAC CTGAAAGCCA TCCAGGACAT GTTCCCAAC ATGGACCAGG
851 AGGTGATCCG CTCGTGCTG GAAGCCAGC GAGGGAACAA GGATGCCGCC
901 ATCAACTCCC TGCTGCAGAT GGGGAGGAG CCATAGAGCC TCTGCCTCGA
951 TGCCGTTTTG CCCCCTCTT TTGGACACGC CGACCCGCGC CTCCTCAAGG
1001 AATGCTGTCC CAACAAGATT CCCGTGAAAG AGCACCCGTG TCGCCCCCTC
1051 CCGTGGACTT CTGTGCCGCC CCGTCCACAC CTGTTCTTGG GTGCATGTGG
1101 GTTTTCTGTT CCTGGCGGTC CAGGACGGGG CGGGGGCTCC CCTCCATCT
1151 CGTCTGGGA GGTCTCAGCG CGCTCTCCTG TCCTGGGAC GTGCGTCTCT
1201 CCTTCTCATG CCGTTCTGGA AAATGCTCTT GCTGTAGAGA GCAGCTGCTT
1251 CTGCCAGGGT GTTGGAGGTG GTGGAGCGCC TTCCGATTCC ATTATGGCA
1301 TTTTGTGATG TGATGTAATT GGAATAGAGC TGTGATTTA AGGCACACAC
1351 AATCCCTCAC ACTGTGGGTT TTTTITAGAA CTTCACAGAC GAAACTCAC
1401 GCCCTTGCCC TAACGCGCTT TGCTGTGAGC CTGGCCCTTG CCCAGGGCTT
1451 GGGCTGTGGT AGCTGAGCAG CTCTCTGTGG ATGGTGTGGG GCCGCGCTCT
1501 GGCCTGGCTC ACCTGGCCAC TGTCCAGCCA GCCTTGTGAC AGACTCCGGC
1551 CTGAAGCCAG AATGAACCCA CACCTGGAGT GAGGAAGGGG GCCTGGCACG
1601 GTTGGCCAGG CTCTGCCTGA TTGCCAGCCA GCGGCGATCT GAAGCCGGGT
1651 CCTTCGCCCG CCGGAGGCTG CCGTCCGTCT CTCTGCTGCG GCTCGTGCCA
1701 GCTCCGTGGG TGTCCTCCCA GGGAGCTTCT CTCTCAACA GGCCTTGCGA
1751 GGCTGGGGTG AGAGGTGATA GAGGCAGCAC TGTGCATGAT TCCGAGAGGG
1801 TGTGGTGGCA CTGCCAGCCG ACTGCTGACA GCTTGGGAGC TGCTGTGCCC
1851 AGGACGTGGG TTCAGCGTGG GCGAGGAAAG CCTGGCGAGC GTGGCCCTGT
1901 AAAAGCTTTC TGAGGCGGGA GCGCTCACT TACCTGTGAC TGCTGGGGCG
1951 CTGCGTGTAG CATCTTGGCC TACAGGACAG ATTTTAGGTG ACACCTGGTT
2001 ATGACAGTCA GAAATTTGAG AAGCTTCTCA CAAGTGATGC ACTTTAAATA
2051 ATCTGCATGC CATTGAGACA CCTGCATGTC TGGTGTGTGT GGTTCAGGTG
2101 TCTTGCCGCC GGCCTTCGGA TGTAAACCCA CTGATAACGG ACAGAAAGAG
2151 AATGCCCACA AGTGGGTCTT CTGTGGAAGA TGCAGAAAGG GGAAGTTAGT
2201 GCTTACATTT TAGTCTTTTT CTCCCTCAA AAAATAGGTT AAGTTTCACT
2251 GCCAGCTAGA AAATACTGCT TTCTGCCATC GATTGGGGGT GGTTTTGTG
2301 AAATATACGT TTGATAAATA TTTATTTTTG TAACTTGAA GTGTGTGGTG
2351 GCGGTGGGGG AGGACATGC TGGCAGCAGG CGCCTTCTTC AGCTGTGGGT
2401 CCTTAAAGGCC TTTGATCCTT TGAAGAAGAA AGACATGGTA TTTGTTACGC
2451 AGACGCCGAC CACTCAGACG GAGGGGCCCC TGGGATTCCC TGTCTCAGAT
2501 GGCTGGTCT TACGCTGTG TAGATTTCTT CTCCATTGGG AATGAAGGTG
2551 TCAGGCGGGA CTGGAACGTT CTAGATGGTA TGTCCGTGA TATTAACAAC
2601 TCTAACCCAG GACAGACCAC AAGCCACACT CAGAGGCTC ACTGTGCTGG
```

```

2651 GGGCTTCGGT GTCCAGGCGC CCAGGTGTGG CCACCAGCAC CGGTTTCTGC
2701 CTTCGCGTTG CTGGGGTGCA GTGAGACTGC CACACGCGTG CACATGTGGC
2751 TCTGTGGGTG TCTCCTAGAG AGGACGTGGC CCCTGCTGCC AGCCCTTGAG
2801 CAGCCCGTGT GGGGGCCCGA GGGACCCACA CAGTGGGGGC CAGCCTCGCT
2851 GGAGGGAGAG CAACCCCTTG CCGATGACCA CGCTTGCCGC CATCTCTTAG
2901 TTTTCTTTT CACAAGCGCT TTATTTTTT AATAGACAAA TCACATTTTG
2951 CAAGGCCTTT AATTAAATAA GATTCTTCTT TCCTTCATT TATGCTTTAT
3001 TTCCTGTTT AAGGCTTACT GTAGAAGTGG CTTACTGTAG AAGCAGCTTG
3051 CTGAGCCCTT CCGAGCGGTC CCCAGAATTA GCTGTTTAC AACCCCCACC
3101 CTCCCCCGCC CCGCCTGTG TCAGGTGTGG ATGAGTCTGT CACACTCAGA
3151 AGGACAGGCT TGTCTGCCAG CTCACAAGGG GAGGCTGCAG TGGGTTTGGG
3201 AGCTGGGTTT AGGCCCTGGT TGTCTGAGGG CCCAGGCCCT GCCAGCCTCT
3251 GCTGCTCCTG CTCTGGGTT TGAAGATGCA GGCCGATCGC CAGCTCCGTG
3301 GCAGCGGTCA CTAAGGACAG CTTGACTGTG CCATCTTGA GCCTCAGGCG
3351 GGGCTCCGGA GATAGAAGAC AGGTCGCCGG AGGCTCCCC TCCTCTCCTC
3401 TCCCTCTGTC AGATGCTCCC TGGGGCGTAC CTTGCAGGCT GCCAGGCAGG
3451 AGTGGTCTCA GAACGTGCGC TTCTGATTAT TTTACTGGGG TCCATTGTCC
3501 AGATTTTCTT TTGATTGTAA AATATATTTT TACTTTTAT TCTTCTAATT
3551 TAATAAATGA TCCATATAAA AATAGAGAAA TAAAGTCCTT TAAGGGAAGG
3601 TTTAAAAAAA AAAA

```

## BLAST Results

No BLAST result

## Medline entries

No Medline entry

## Peptide information for frame 1

ORF from 112 bp to 933 bp; peptide length: 274  
 Category: similarity to unknown protein  
 Classification: no clue

```

1 MATTVSTQRG PVYIGELPQD FLRITPTQQQ RQVQLDAQAA QQLQYGGAVG
51 TVGRLNITVV QAKLAKNYGM TRMDPYCRLR LGYAVYETPT AHNGAKNPRW
101 NKVIHCTVPP GVDSFYLEIF DERAFSMDDR IAWTHITPE SLRQGVKVEDK
151 WYSLSGRQGD DKEGMINLVM SYALLPAAMV MPQPVVVLM TVYQQGVGVV
201 PITGMPAVCS PGMVPVALPP AAVNAQPRCS EEDLKAIQDM FPNMDQEVIR
251 SVLEAQRGNK DAAINSLQGM GEEP

```

## BLASTP hits

No BLASTP hits available

## Alert BLASTP hits for DKF2phtes3\_2dl5, frame 1

TREMBL:CEF25H2\_1 gene: "F25H2.1"; *Caenorhabditis elegans* cosmid F25H2,  
 N = 1, Score = 385, P = 1.1e-35

>TREMBL:CEF25H2\_1 gene: "F25H2.1"; *Caenorhabditis elegans* cosmid F25H2  
 Length = 457

## HSPs:

Score = 385 (57.8 bits), Expect = 1.1e-35, P = 1.1e-35  
 Identities = 77/182 (42%), Positives = 118/182 (64%)

```

Query: 4 TVSTQRGPVYIGELPQDFLRIT-PTQQQRQVQLDAQAAQQLQYGGAVGTGRLNITVVQA 62
      TV+ +R V +GELP FLR+ P QQ + ++ Q + + + T GRL++T+++A
Sbjct: 5 TVAERRRQVLVGELPPHFLRLAVPIQQTAEPEI-VQP-RMVSVFPP-NTRGRLSVTILEA 61

Query: 63 KLAKNYGMTRMDPYCRLRLGYAVYETPTAHNGAKNPRWNKVIHCTVPPGVDSFYLEIFDE 122
      L KNYG+ RMDPYCR+R+G ++T A N + P WN+ ++ +P V+S Y++IFDE
Sbjct: 62 NLVKNYGLVRMDPYCRVRVGNVEFDTNVAANAGRAPTWNRTLNAYLPMNVESIYIIFDE 121

Query: 123 RAFSMDDRIAWTHITIPESLRQGVKVEDKWYSLSGRQGGDDKEGMINLVMSYAL--LPAAMV 180
      +AF D+ IAW HI +P ++ G D+++ LSG+QG+ KEGMI+L S+A LP
Sbjct: 122 KAFGPDEVIWAHIMLPLAIFNGONIDEYFQLSGQQGEGKEGMIHLHFSFAPIDLPLQQA 181

```



Query: 181 MPPQP 185

P +P

Sbjct: 182 APAEP 186

Score = 92 (13.8 bits), Expect = 1.8e-01, P = 1.7e-01  
Identities = 26/68 (38%), Positives = 38/68 (55%)

Query: 194 QQGVGYVPITGMPAVCSPGMVVPV--ALP--PAAVNAQPRCSEEDLKAIQDMFPMNDQEV 249

QQG G + + +P +P+ A P PA +EED K IQ+MFP +D+EVI

Sbjct: 156 QQGEGKEGMIHLHFSFAPIDLPLQQAAPAEPAAPLPVEITEEDTKEIQEMFPIVDKEVI 215

Query: 250 RSVLEAQR 257

+ +LE +R

Sbjct: 216 KCILEERR 223

Pedant information for DKFZphtes3\_2d15, frame 1

Report for DKFZphtes3\_2d15.1

[LENGTH] 274  
[MW] 30281.97  
[pI] 5.68  
[HOMOL] TREMBL:CEF25H2\_1 gene: "F25H2.1"; Caenorhabditis elegans cosmid F25H2 4e-36

[PFAM] C2 domain  
[KW] Alpha\_Beta  
[KW] LOW\_COMPLEXITY 16.42 %

SEQ MATTVSTQRGPVYIGELPQDFLRITPTQQQRQVQLDAQAAQQLQYGGAVGTGRLNITVV  
SEG .....XXXXXXXXXXXXXXXXX.....  
PRD cccccccccceccccccccccccchhhhhhhhhhhhhhhhhcccccccececeeh

SEQ QAKLAKNYGMTRMDPYCRLRLGYAVYETPTAHNGAKNPRWNKVIHCTVPPGVDSFYLEIF  
SEG .....  
PRD hhhhhhhccccccccchhhheeeeecccccccccccccecccccccccecececece

SEQ DERAFSMDDRIAWTHITIPESLRQGVKEDKWYSLSGRQGGDDKEGMINLVMSYALLPAAMV  
SEG .....XXXXXXXXX  
PRD ccccccccccecccccccccccccccccecccccccccccccececececehhhhhhhhc

SEQ MPPQPVVLMPTVYQQGVGYVPITGMPAVCSPGMVVPVALPPAAVNAQPRCSEEDLKAIQDM  
SEG xxxxxxxxxxxxxxxxxxxxxxxxx  
PRD ccccccecccccccccccccccccecccccccccccccececcccchhhhhhhhhc

SEQ FPNMDQEVIRSVLEAQRGNKDAAINSLLQMGEEP  
SEG .....  
PRD cccccchhhhhhhhhccccchhhhhhhhhccc

(No Prosite data available for DKFZphtes3\_2d15.1)

Pfam for DKFZphtes3\_2d15.1

HMM\_NAME C2 domain  
HMM \*LTVrIIeARNLWkMDMnGfSDPYVKVdMdPdpkDtkKWktKiWNNGLN  
L+++++A+ + + M+ DPY+++ + + +T T +N N  
Query 55 LNITVVQAKLAKNYGMT-RMDPYCRLRLGYAVY-----ETPTAHNGAKN 97  
HMM PVWNEEeFvFedIPyPdIqrkMLRFaVWDWDRFSRBDFIGHCi\*  
P+WN + +P + + ++++D+ FS +D I+ +  
Query 98 PRWN-KVIHCT-VPPGVDSF---YLEIFDERAFSMDDRIAWTH 135

DKFZphtes3\_2e12  
-----

group: Transcription Factors

DKFZphtes3\_2e12 encodes a novel 849 amino acid protein with similarity to Zinc finger proteins.

The new protein is a putative transcription factor with three C2H2 zinc fingers. Additionally, a cytochrome C family heme-binding site signature is present in the protein, which is only found in cytochrom C related proteins.

The new protein can find application in modulating/blocking the expression of genes controlled by this transcription factor.

similarity to finger proteins

complete cDNA, complete cds, 5 EST hits

Sequenced by EMBL

Locus: unknown

Insert length: 3205 bp

Poly A stretch at pos. 3192, polyadenylation signal at pos. 3171

```
1  GGCACGGCCG GGTCTGGCT GGCCAAACGA GGCTCGCGGA AGCAGCAGCC
51  GCCCCCTGAC CGCAGCTGGA TTTGAAGAT TGATCCAAGG GACTGTATTA
101 ATTTCAGGAA TTGATTTGAA AGACACTGGC TCTGCCACTT AACAGCCATG
151 TAACCTTGGA TATGGAAGAA AGTAGCAGTG TTGCCATGTT GGTGCCAGAT
201 ATTGGGGAAC AGGAAGCTAT ACTGACTGCT GAAAGTATCA TCAGTCCTTC
251 ATTGGAAATT GATGAACAAA GAAAACTAA ACCAGATCCA TTAATCCATG
301 TTATCCAGAA GTTAAGCAAG ATAGAAAAAT GAAAAGTCAC AAAAATGTCT
351 TTTAATTGGG AAGAAACGCC CACGTTCAGG TGCTGCAACA CACTCTCTTG
401 AAACCAAGA ACTTTGTGAG ATTCCGGCTA AAGTAATCCA GTCACCTGCT
451 GCTGATACTA GAAGGGCTGA GATGTCACAA ACAAAATTTA CCCCTGACAC
501 TCTTGCCAG AATGAAGGGA AGGCTATGTC TTATCAGTGT AGCCTTTGTA
551 AGTTTCTATC ATCATCCTTT TCCGTGTTAA AAGATCATAT TAAGCAACAT
601 GGTCAGCAAA ATGAAGTGAT ACTGATGTGC TCAGAGTGCC ATATTACATC
651 TAGAAGCCAG GAGGAACCTG AAGCCACGTT GGTGAATGAC CATGACAATG
701 ATGCCAATAT CCACACCCAA TCCAAAGCCC AACAGTGCGT AAGCCCCCTC
751 AGCTCTTTGT GTCGAAAAC CACAGAAAGA AATGAAACCA TTCCAGATAT
801 CCCAGTAAGT GTGGACAATC TACAGACTCA TACTGTCCAA ACTGCATCTG
851 TGGCAGAAAT GGGTAGGAGG AAATGGTATG CATACGAACA GTACGGCATG
901 TATCAGATGCT TGTTTGTAG TTATACTTGT GGCCAGCAGA GAATGTTGAA
951 AACACACGCT TGGAAACATG CTGGGGAGGT TGATTGCTCC TATCCAACTC
1001 TTGAAAATGA AAATGAACCC CTAGGCCCTG TGGATTCTTC AGCAGCTGCT
1051 GCGCTGGTG GGGTCGATGC AGTCGTCTAT GCTATTGGAG AGAGTGAACT
1101 GAGTATCCAC AATGGGCCAT CAGTGCAAGT GCAGATTGCG AGCTCAGAAC
1151 AGTTATCATC TTCATCTCCT TTAGAACAGA GTGCAGAAAG AGGAGTACAC
1201 CTAAGTCAGT CAGTTACCTT GGACCCCAAT GAGGAAGAAA TGCTAGAAGT
1251 GATTTCTGAT GCAGAGGAGA ATCTGATTCC TGATAGCCTG CTTACATCAG
1301 CACAGAAAAT CATCAGCAGC AGCCCCAATA AAAAAGGGCA TGTTAACGTG
1351 ATAGTGGAGC GATTGCCAAG TGCTGAAGAA ACCCTTTCAC AGAAGCGCTT
1401 CCTCATGAAC ACTGAAATGG AAGAAGGGAA GGACCTGAGC CTGACAGAAG
1451 CTCAGATTGG GCGCGAAGGA ATGGATGATG TTTATCGTGC TGATAAATGT
1501 ACTGTTGATA TTGGGGGATT GATCATAGGC TGGAGCAGTT CAGAGAAAAA
1551 AGACGAGTTA ATGAATAAAG GCCTGGCTAC TGATGAGAAT GCCCACCAG
1601 GCCGGAGAAG GACAAATTCT GAGTCTCTTC GATTACACTC ATTAGCTGCA
1651 GAAGCCCTTG TCACAAATGCC TATAAGAGCT GCAGAGTTGA CAAGAGCCAA
1701 CCTGGGGCAC TATGGAGATA TAAACCTTTT AGATCCAGAT ACTAGTCAAA
1751 GGCAAGTAGA TAGTACATTG GCAGCGTACT CAAAAATGAT GTCGCCACTT
1801 AAAAATCTTT CAGATGGATT AACTAGTCTT AACCAAAGCA ACTCCACCTT
1851 GGTAGCACTC CCAGAGGGTA GGCAGGAATT GTCAGATGGG CAGGTTAAGA
1901 CAGGCATCAG CATGTCCTTA CTCACCGTCA TTGAAAAATT GAGAGRAAGG
1951 ACAGACCAAA ACGCTTCAGA CGATGACATT TTGAAAGAGT TGCAGGACAA
2001 CCCCCAGTGC CAACCAACAA GCGATACAAG TTTGTCCGGA AACAAATGTG
2051 TGGAAATACAT CCCGAATGCT GAACGACCTT ACCGTTGCCG CCGTGTTCAC
2101 TACACAAGTG GCAACAAGGG CTACATCAAG CAGCACTTAC GAGTCCATCG
2151 ACAGGACAGC CTTATCAGT GTCCTATCTG CGAGCACATA GCGGACAACA
2201 GCAAAGATTT GGAGAGTCAC ATGATCCACC ACTGTAAGAC AAGAATATAC
2251 CAGTGCAAGC AGTGTGAAGA ATCCTTCCAT TATAAGAGTC AATTGAGGAA
2301 CCATGAGAGA GAACAGCACA GTCTTCCAGA TACCTTGTC AATGCAACTT
2351 CTAATGAGCC AAGAATTTCC AGTGATACAG CTGATGGAAA ATGTGTCAG
2401 GAAGGGAATA AGTCTTCAGT CCAGAAACAA TATAGATGTG ATGTGTGTGA
2451 TTATACAAGT ACAACATATG TTGGTGTGAG AAACCCACAG CGAATCCATA
2501 ACTCTGATAA GCCGTACAGA TGCTCTCTGT GTGGGTATGT GTGTAGCCAT
2551 CCTCCTCTCT TGAAGTCTCA TATGTGGAAG CATGCAAGTG ACCAAAAATTA
```

```

2601 CAACTACGAA CAAGTAAACA AGGCTATTAA CGACGCGATT TCACAAAGTG
2651 GCAGAGTTCT GGGGAAATCC CCTGGAAAGA CTCAATTAAA GAGCAGTGAA
2701 GAGAGTGCAG ATCCCGTCAC TGGAACTTCG GAAATGCAG TGTCTCTTC
2751 AGAACTGATG TCCCAGACTC CCAGTGAAGT TCTGGGTACC AACGAGAATG
2801 AGAACTGAG CCCTACAAGT AATACCTCAT ATAGTTTAGA AAAAACTCC
2851 AGTCTGGCCC CTCCTAGCAT GGAGTACTGC GTTTTACTCT TCTGCTGTG
2901 TATTTGTGGT TTTGAATCAA CCAGCAAAGA AAACCTCTTG GATCATATGA
2951 AAGAGCACGA GGGTGAAATT GTAAACATCA TCCTGAATAA GGACCAACAAT
3001 ACAGCTCTAA ACACAAATTA GGTGGAATAA TGACTCGAGC AGGAAAGCAG
3051 TAGAAGAGGA TTCCTTCACC ACAGTTTCAC CTTTACGCTG TCAGACAACT
3101 TCCTGCCACA GAAGAAGTCG TTGATGTGAT TTTTGAGGAA ATGACAGATG
3151 TGACTTTGGA ACCAAACTTG TAATAAAGG AATTCCAAAT GGAACAAAAA
3201 AAAAA

```

# BLAST Results

No BLAST result

# Medline entries

90301500:  
Cloning and sequencing of a zinc finger cDNA expressed in mouse testis.

92310982:  
Zfp-37, a new murine zinc finger encoding gene, is expressed in a developmentally regulated pattern in the male germ line.

# Peptide information for frame 1

ORF from 472 bp to 3018 bp; peptide length: 849  
Category: similarity to known protein

```

1 MSQTNFTPDT LAQNEGKAMS YQCSLCKFLS SSFVSLKDHI KQHGOQNEVI
51 LMCSECHITS RSQEELEAHV VNDHDNDANI HTQSKAQQCV SPSSSLCRKT
101 TERNETIPDI PVSVDNLQTH TVQTASVAEM GRRKWAYAYE YGMYRCLFCS
151 YTCGQQRMLK THAWKHAGEV DCSYPIFENE NEPLGLLDSS AAAAPGGVDA
201 VVIAIGESSEL SIHNGPSVQV QICSSEQLSS SSPLEQSAER GVHLSQSVTL
251 DPNEEEMLEV ISDAEENLIP DSLTSAQKI ISSSPNKKGH VNVIVERLPS
301 AEETLSQKRF LMNTEEMEGK DLSLTEAQIG REGMDDVYRA DKCTVDIGGL
351 IIGWSSSEKK DELMNKGLAT DENAPPGRRR TNSESLRLHS LAEALVTMP
401 IRAAELTRAN LGHYGDINLL DPDTSORQVD STLAAYSKMM SPLKNSSDGL
451 TSLNQSNTSL VALPEGRQEL SDGOVKTGIS MSLLTVIEKL RERTDQNASD
501 DDILKELQDN AQCPQNSDTS LSGNNVVEYI PNAERPYPYR LCHYTSGNKG
551 YIKQHLRVHR QRQPYQCPIC EHIADNSKDL ESHMIHCKT RIYQCKQCEE
601 SFHYKSLRN HEREQHSPLD TSLIATSNEP RISSDTADGK CVQEGNKKSV
651 QKQYRCDVCD YTSTTYVGVN NHRRIHNSDK PYRCSLCGYV CSHPPSLKSH
701 MWKHASQNY NYEQVNKAIN DAISQSGRVL GKSPGKTQLK SSEESADPVT
751 GSSENAVSSS ELMSQTPSEV LGTNEKLS PTSNTSYSLE KISSLAPPSM
801 EYCVLLFCCC ICGFESTSKE NLLDHMKHE GEIVNIILNK DHNTALNTN

```

# BLASTP hits

Entry S10245 from database PIR:  
finger protein, testis - mouse  
Score = 265, P = 8.4e-23, identities = 61/205, positives = 91/205

Entry S22954 from database PIR:  
finger protein zfp-37 - mouse  
Score = 265, P = 9.1e-22, identities = 61/205, positives = 91/205

Entry AF031657.1 from database TREMBL:  
gene: "Zfp94"; product: "zinc-finger protein 94"; Rattus norvegicus  
zinc-finger protein 94 (Zfp94) gene, partial cds.  
Score = 243, P = 1.6e-21, identities = 57/190, positives = 85/190

Alert BLASTP hits for DKF2phtes3\_2e12, frame 1

No Alert BLASTP hits found

Pedant information for DKFZphtes3\_2e12, frame 1

Report for DKFZphtes3\_2e12.1

[LENGTH] 849  
 [MW] 94325.42  
 [pI] 5.47  
 [HOMOL] PIR:A54661 zinc finger protein 2NF41 - human (fragment) 2e-22  
 [FUNCAT] 04.05.01.04 transcriptional control [S. cerevisiae, YJL056c] 3e-09  
 [FUNCAT] 30.10 nuclear organization [S. cerevisiae, YJL056c] 3e-09  
 [FUNCAT] 04.03.01 trna synthesis [S. cerevisiae, YPR186c PZF1 - TFIIIA] 1e-07  
 [FUNCAT] 04.01.01 rrna synthesis [S. cerevisiae, YPR186c PZF1 - TFIIIA] 1e-07  
 [FUNCAT] 04.99 other transcription activities [S. cerevisiae, YOR113w] 4e-07  
 [FUNCAT] 01.05.04 regulation of carbohydrate utilization [S. cerevisiae, YGL209w] 2e-04  
 [FUNCAT] 13.04 homeostasis of other ions [S. cerevisiae, YNL027w] 2e-04  
 [FUNCAT] 11.01 stress response [S. cerevisiae, YMR037c] 3e-04  
 [BLOCKS] BL00028 Zinc finger, C2H2 type, domain proteins  
 [SCOP] dlmeyg\_9.6.1.1.1 a designed zinc finger protein [syntheti 8e-06  
 [PIRKW] nucleus 8e-18  
 [PIRKW] RNA binding 5e-13  
 [PIRKW] duplication 7e-13  
 [PIRKW] tandem repeat 1e-21  
 [PIRKW] spermatogenesis 6e-16  
 [PIRKW] zinc 9e-21  
 [PIRKW] zinc finger 1e-21  
 [PIRKW] DNA binding 1e-21  
 [PIRKW] metal binding 3e-15  
 [PIRKW] phosphoprotein 5e-13  
 [PIRKW] leucine zipper 1e-13  
 [PIRKW] alternative splicing 6e-18  
 [PIRKW] eye lens 2e-16  
 [PIRKW] oocyte 1e-12  
 [PIRKW] transcription factor 6e-18  
 [PIRKW] segmentation 7e-13  
 [PIRKW] embryo 1e-12  
 [PIRKW] transcription regulation 2e-19  
 [PIRKW] homeobox 2e-08  
 [SUPFAM] POZ domain homology 7e-15  
 [SUPFAM] transcription factor Krueppel 7e-13  
 [SUPFAM] zinc finger protein ZFP-36 1e-21  
 [SUPFAM] homeobox homology 2e-08  
 [SUPFAM] unassigned homeobox proteins 2e-08  
 [PROSITE] CYTOCHROME\_C 1  
 [PROSITE] MYRISTYL 10  
 [PROSITE] ZINC\_FINGER\_C2H2 3  
 [PROSITE] AMIDATION 2  
 [PROSITE] CAMP\_PHOSPHO\_SITE 2  
 [PROSITE] CK2\_PHOSPHO\_SITE 18  
 [PROSITE] TYR\_PHOSPHO\_SITE 3  
 [PROSITE] PKC\_PHOSPHO\_SITE 10  
 [PROSITE] ASN\_GLYCOSYLATION 7  
 [PFAM] Zinc finger, C2H2 type  
 [KW] Irregular  
 [KW] 3D  
 [KW] LOW\_COMPLEXITY 5.65 %

SEQ MSQTNFTPDTLAQNEGKAMSYQCSLCKFLSSSFVLKDHQKQHQNEVILMCSECHITS  
 SEG .....XXXXXXXXXXXXXXXXXXXX  
 lmeyF .....  
 SEQ RSQEELEAHVVNDHDNDANIHTQSKAQCVSPSSSLCRKKTTERNETIPDIPVSVNQLQTH  
 SEG .....  
 lmeyF .....  
 SEQ TVQTASVAEMGRKRWYAYEQYGMRYCLFCSYTCGQQRMLKTHAWKHAGEVDCSYPIFENE  
 SEG .....  
 lmeyF .....  
 SEQ NEPLGLLDSSAAAAAPGGVDVAVIAIGESLSIHNGPSVQVQCSSSEQLSSSSPLEQSAER  
 SEG .....XXXXXXXXXXXXXXXXXXXX  
 lmeyF .....  
 SEQ GVHLSQSVTLDPNEEMLEVISDAEENLIPDSLLTSAQKIISSPNKKGHVNVIVERLPS  
 SEG .....  
 lmeyF .....

```

SEQ      AEETLSQKRFLMNTMEMEKGDLSTEAQIGREGMDDVYRADKCTVDIGGLIIGWSSEKK
SEG      .....
lmeYF    .....

SEQ      DELMNKGLATDENAPPGRRTNSESRLHSLAAEALVTMPIRAAELTRANLGHYGDINLL
SEG      .....
lmeYF    .....

SEQ      DPDTSQRVQDSTLAAYSKMMSPLKNSSDGLTSLNQSNTLVALPEGRQELSDGQVKTGIS
SEG      .....
lmeYF    .....

SEQ      MSLLTVIEKLRERTDQNASDDDLKELQDNAQCQPNSDTSLSGNNVVEYIPNAERP YRCR
SEG      .....
lmeYF    .....TTTEETT

SEQ      LCHYTSNGKGYIKQHLRVHRQRQPYQCPICEHIADNSKDLESHMIHCKTRIYQCKQCEE
SEG      .....
lmeYF    TTTCEETTHHHHHHHHHHTTCCCEETTTTTEECCHHHHHHHHHHHHCCCCCEETTTTE

SEQ      SFHYKSQLRNHEREQHSLPDTLSIATSNEPRISSDTADGKCVQEGNKSSVQKQYRCDVCD
SEG      .....
lmeYF    EECCHHHHHHHHHHHC.....

SEQ      YTSTTYVGVNRHRIHNSDKPYRCSLCGYVCSHPPSLKSHMKHASDQNYNEQVNKAIN
SEG      .....
lmeYF    .....

SEQ      DAISQSGRVLGKSPGKTQLKSSEESADPVTGSSENAVSSSELSQTPSEVLGTNENEKLS
SEG      .....
lmeYF    .....

SEQ      PTSNTSYSLEKISSLAPPSMEYCVLLFCCCICGFESTSKENLLDHMKHEGEIVNIILNK
SEG      .....
lmeYF    .....

SEQ      DHNTALNTN
SEG      .....
lmeYF    .....

```

## Prosites for DKFZphtes3\_2e12.1

PS000001	104->108	ASN_GLYCOSYLATION	PDOC00001
PS000001	445->449	ASN_GLYCOSYLATION	PDOC00001
PS000001	454->458	ASN_GLYCOSYLATION	PDOC00001
PS000001	457->461	ASN_GLYCOSYLATION	PDOC00001
PS000001	497->501	ASN_GLYCOSYLATION	PDOC00001
PS000001	646->650	ASN_GLYCOSYLATION	PDOC00001
PS000001	784->788	ASN_GLYCOSYLATION	PDOC00001
PS000004	98->102	CAMP_PHOSPHO_SITE	PDOC00004
PS000004	378->382	CAMP_PHOSPHO_SITE	PDOC00004
PS000005	59->62	PKC_PHOSPHO_SITE	PDOC00005
PS000005	101->104	PKC_PHOSPHO_SITE	PDOC00005
PS000005	306->309	PKC_PHOSPHO_SITE	PDOC00005
PS000005	357->360	PKC_PHOSPHO_SITE	PDOC00005
PS000005	385->388	PKC_PHOSPHO_SITE	PDOC00005
PS000005	425->428	PKC_PHOSPHO_SITE	PDOC00005
PS000005	678->681	PKC_PHOSPHO_SITE	PDOC00005
PS000005	696->699	PKC_PHOSPHO_SITE	PDOC00005
PS000005	726->729	PKC_PHOSPHO_SITE	PDOC00005
PS000005	817->820	PKC_PHOSPHO_SITE	PDOC00005
PS000006	62->66	CK2_PHOSPHO_SITE	PDOC00006
PS000006	106->110	CK2_PHOSPHO_SITE	PDOC00006
PS000006	126->130	CK2_PHOSPHO_SITE	PDOC00006
PS000006	232->236	CK2_PHOSPHO_SITE	PDOC00006
PS000006	262->266	CK2_PHOSPHO_SITE	PDOC00006
PS000006	300->304	CK2_PHOSPHO_SITE	PDOC00006
PS000006	314->318	CK2_PHOSPHO_SITE	PDOC00006
PS000006	323->327	CK2_PHOSPHO_SITE	PDOC00006
PS000006	355->359	CK2_PHOSPHO_SITE	PDOC00006
PS000006	381->385	CK2_PHOSPHO_SITE	PDOC00006
PS000006	485->489	CK2_PHOSPHO_SITE	PDOC00006
PS000006	499->503	CK2_PHOSPHO_SITE	PDOC00006
PS000006	617->621	CK2_PHOSPHO_SITE	PDOC00006
PS000006	626->630	CK2_PHOSPHO_SITE	PDOC00006
PS000006	741->745	CK2_PHOSPHO_SITE	PDOC00006
PS000006	758->762	CK2_PHOSPHO_SITE	PDOC00006
PS000006	766->770	CK2_PHOSPHO_SITE	PDOC00006
PS000006	817->821	CK2_PHOSPHO_SITE	PDOC00006

PS00007	331->339	TYR_PHOSPHO_SITE	PDOC00007
PS00007	703->711	TYR_PHOSPHO_SITE	PDOC00007
PS00007	596->605	TYR_PHOSPHO_SITE	PDOC00007
PS00008	142->148	MYRISTYL	PDOC00008
PS00008	185->191	MYRISTYL	PDOC00008
PS00008	196->202	MYRISTYL	PDOC00008
PS00008	241->247	MYRISTYL	PDOC00008
PS00008	349->355	MYRISTYL	PDOC00008
PS00008	473->479	MYRISTYL	PDOC00008
PS00008	478->484	MYRISTYL	PDOC00008
PS00008	645->651	MYRISTYL	PDOC00008
PS00008	751->757	MYRISTYL	PDOC00008
PS00008	772->778	MYRISTYL	PDOC00008
PS00009	130->134	AMIDATION	PDOC00009
PS00009	376->380	AMIDATION	PDOC00009
PS00028	146->167	ZINC_FINGER_C2H2	PDOC00028
PS00028	684->705	ZINC_FINGER_C2H2	PDOC00028
PS00028	595->617	ZINC_FINGER_C2H2	PDOC00028
PS00190	53->59	CYTCHROME_C	PDOC00169

## Pfam for DKFZphtes3\_2e12.1

HMM\_NAME Zinc finger, C2H2 type

HMM \*CpwPDCgKtFrrwsNlrRHMRT.H\*  
C++ C+ T R+++L++H H

Query 53 CSE--CHITSRSQEELAHVVN-DH 74

23.25 (bits) f: 539 t: 559 Target: dkfzphes3\_2e12.1 similarity to finger proteins  
Alignment to HMM consensus:  
Query \*CpwPDCgKtFrrwsNlrRHMRT.H\*  
C C++T ++ ++H+R+H  
dkfzphes3 539 CRL--CHYTSNGYIKQHLRVH 559

Query f: 567 t: 587 Target: dkfzphes3\_2e12.1 similarity to finger proteins  
Alignment to HMM consensus:  
HMM \*CpwPDCgKtFrrwsNlrRHMRT.H\*  
CP+ C+ ++ +L+ HM+ H  
Query 567 CPI--CEHIADNSKDLESHMIHH 587

33.47 (bits) f: 595 t: 616 Target: dkfzphes3\_2e12.1 similarity to finger proteins  
Alignment to HMM consensus:  
Query \*CpwPDCgKtFrrwsNlrRHMRT.H\*  
C+ C+++F ++S+LR+H R H  
dkfzphes3 595 CKQ--CEESFHYKSQRLNHERE-QH 616

Query f: 656 t: 676 Target: dkfzphes3\_2e12.1 similarity to finger proteins  
Alignment to HMM consensus:  
HMM \*CpwPDCgKtFrrwsNlrRHMRT.H\*  
C++ C++T ++ R+H+R+H  
Query 656 CDV--CDYTSTTYVGVNRHRIH 676

24.53 (bits) f: 684 t: 704 Target: dkfzphes3\_2e12.1 similarity to finger proteins  
Alignment to HMM consensus:  
Query \*CpwPDCgKtFrrwsNlrRHMRT.H\*  
C+ CG++ +++ +L+ HM H  
dkfzphes3 684 CSL--CGYVCSHPPLKSHMWKH 704

Query f: 809 t: 829 Target: dkfzphes3\_2e12.1 similarity to finger proteins  
Alignment to HMM consensus:  
HMM \*CpwPDCgKtFrrwsNlrRHMRT.H\*  
C+ CG +++NL HM+ H  
Query 809 CCI--CGFESTSKENLLDHMKEH 829

DKFZphtes3\_2f14  
-----

group: testes derived

DKFZphtes3\_2f14 encodes a novel 129 amino acid protein with very weak similarity to human omega protein.

No informative BLAST results; No predictive prosite, pfam or SCOP motif.

The new protein can find application in studying the expression profile of testis-specific genes.

weak similarity to omega protein

complete cDNA, complete cds, 1 EST hit

Sequenced by EMBL

Locus: unknown

Insert length: 2353 bp  
Poly A stretch at pos. 2341, no polyadenylation signal found

```
1 GCAGATTCTC CAGGCCAGC ATCTGCCTCA CCGTGGCCCC CCACAAGCCA
51 AGCGCCTGCC TTTCAGCAGC CTCTACACAC CCAGCTCCTG CCACCCAATG
101 GCTCTTTAGG CCAAGCTCAT ACCTCAGCAT GATTTTCCA GGCCCAACTT
151 TTGTCTCATG GCAACCTTCC CTGGCCAAGT TTCCACCTAT TTCCTGGCAG
201 CCTGGACAGG CCCAGGTCCT GCCACACACT GGCTCTCTA CGCCACAGCTC
251 ATGCCTCACA GTGGCCTCTC CAGGCCAGC TCCTGTCCCG GGACATCATC
301 TCCAGGCCCA AACTTCCTC AAGTCGGCCT CTCCAGGCC AGTTGCTGCC
351 TCCCGGCATT CTCTCCAGGC CTAGCTCTTC CTCTGGCTG TATCTACAAG
401 ACCAACTCCT GCCTCACAA ACCTTTTAT GGCTCAGCTC CTGCCCAACT
451 ACTGCCGGCC TTGTAGGCC CAAAACCTCC TCAAGTCAAG CTCTTTAGGC
501 CCACCTTCTG CTTGCAGTG GCCTGTACAG ACCCAGCTCT GGCTTGAGAA
551 CAGCTCTGTC AGGCCCTGCT CTTGCCTCT AGCTCCCTCT CCAGGCCCAT
601 CTCTTGCCCTC ACAGTGGCTT CCGTGGGCCA AGTCCCGCC TGCCTCCAG
651 CAGCCTCAAC AGGCCCTAGCT CCTCCCTCAC AATGGCTTGT TTAGGTCAG
701 TTGATGCCTC TGGCAACCTG TCCAGGCCCA GCTCCTGCCT CACACTGGCC
751 TCTCTAGGCC GAGGTCCTTT CTCATACTGG CCTGTTTAGG CCCAGCTCAT
801 TCCTCTTGTC ATCTCTCCAG GCCCAGCTTT TGCTGTGTTG TGGCCTCTAC
851 CTCACAGTGC ACCTTCCAGT CCCACCTCTT GCCTCACCAT GGCCTCCTCT
901 GACCAGGTTT CTGCCTTTCC GCAGCCTCTA CAGGCCTAGC TGCTGCCTCC
951 CAATGGCCTT TGTAGGCCAC GCTCATGCCT CACTGTGGCC TTTCCAGGCC
1001 TAGCTTTTGC TTTTGGCCA CTCAGGCCCC AGAACTTCCC CCAGTCAGCC
1051 TCTCCAGGCC CAGCTCTTCC TCCAGCCAAC CTCTGCAGGC CCAATCATC
1101 CTCAAATTGG CCTCTTCTT CCCAGCTCCT GCCTCCTGGT GGCCTCTGAA
1151 GACCCAAATC GTCTCCAGT TGGTTTTC AGGCCAGCT CCTGCCTTTT
1201 GGTGGCCTCT CCAGGTGCAA AACTTCTCTC CATCAGCCTG TCCAGGCCCA
1251 GCTCATGCCT CTGGTGGCC TTCTCAGGCC CTGCTTTTGA CTGGTGGGCC
1301 TCTTCAGGCC CAGAACTTGA ACTCAAGTCA GCCTCTCCAG GCCCAGCTCC
1351 TGCTTCTTTA AGGTCTGTAC AGGCCCAGCC TCTACCTCAC AGCGGACTCT
1401 CCACACCCAG CTCTTGCCCTC ACTGTAGCCT CCCCAGTCCA AAACCTCTGC
1451 CTTTTGGCAG CTTCGACAAG CCCAGCTCCT GCCTTTCAAT GACCTCTTTA
1501 GGCCCGCCTC ATTCCTTACA ACGGCCTTTC CAGGCCCAGT TTTTCCCTTT
1551 TGGCGGCCCTC TCCAGGCCCA GAACTTCTC AAGTCGGCCT CTTTAGGCC
1601 AGTTGCTGCC TCCTGGCCTC CTCTGCAGGC CGAGCTCTTC CTCCCTGCTG
1651 TGTCTACAGG CCAACTCCT GCCTCACAA AACCTCCTTG GACTCAGCTT
1701 CTGCCCAGCT CTTGGTGGCC TTTGTAGGCT CAAATTTTC TCAAATCAAG
1751 CTCTCCAGGC CTACTGTGAG CCTCGTGGCA GCCTAAACAG GCCCAGCTCC
1801 TGCTTGACAA TGGCCTCTCC AGGCTTTTCT CTGCCTCGC AGCAGGCTTT
1851 CCAGGCCCAG CTCTTGCCCTC ATGGTGGCCT TCCCGGCCA TGTTCCTATC
1901 TGACTTCTGG CAGCCTCAAC CGGCCCAGCT TCTGCCTCAC ACTGGCCTCT
1951 CTAGGCCCAG CTCTTTTTC ACAGTGGCCT CACTACGCCC ATCTCCTAGC
2001 TCAGATCTGC CTCCAAGAC CCAGCTCCTG TCTCATGGTG GTCTCTCTTA
2051 CACCAGCTCC TGCTCACA TGGCCTCGTC TGGCCCATCT TCTGCCTCAC
2101 AGTGGCCACT CAAGGCCCAT CTTTGGCCTC ATGGTAGCCT CTCTGGTTT
2151 TGCTCTTGCC TCACAGTTGC CTCTTCCAGA TCCAGCTTTA AGCCTTTGAT
2201 GGTCAACAGC ATCAAGGAGC CTAAGCTTTC CCTGGACTCT CATTTGTTCA
2251 CTTTACAGCA GAGTGCCTTA GCAAAAAC TGCTCTTAAC TTGAGAGTGG
2301 ATTTCTGACA AATCGATAGT AAATCTGCC TGTGTGGTTT CAAAAAATAA
2351 AAA
```

BLAST Results  
-----

No BLAST result

## Medline entries

No Medline entry

## Peptide information for frame 2

ORF from 158 bp to 544 bp; peptide length: 129  
Category: similarity to known protein

1 MATFPGQVST YFLAAWTGPG PATHWPLYAQ LMPHSGLSRP SSCPGTSSPG  
51 PKLPQVGLSR PSCCLPAFSP GLALPPGCIY KTNSCLTTTF YGSAPAQLLP  
101 AFVGPQLPQV KLFRTFCCLA VACTDPALA

## BLASTP hits

Entry I70697 from database PIR:  
omega protein - human (fragment)  
Score = 79, P = 2.8e-03, identities = 32/94, positives = 38/94

## Alert BLASTP hits for DKFZphtes3\_2f14, frame 2

No Alert BLASTP hits found

## Pedant information for DKFZphtes3\_2f14, frame 2

## Report for DKFZphtes3\_2f14.2

[LENGTH] 129  
[MW] 13421.76  
[pI] 9.14  
[PROSITE] MYRISTYL 2  
[KW] Irregular  
[KW] LOW\_COMPLEXITY 10.85 %

SEQ MATFPGQVSTYFLAAWTGPGPATHWPLYAQLMPHSGLSRPSSCPGTSSPGPKLPQVGLSR  
SEG .....XXXXXXXXXXXXXXXXX.....  
PRD cccccccceehhhhhcc

SEQ PSCCLPAFSPGLALPPGCIYKTNSCLTTTFYGSAPAQLLPAFVGPQLPQVKLFRTFCCLA  
SEG .....  
PRD ccccccccccccccccccccccecccccccccccccccccccccccccccccccccccc

SEQ VACTDPALA  
SEG .....  
PRD cccccccc

## Prosite for DKFZphtes3\_2f14.2

PS00008 6->12 MYRISTYL PDOC00008  
PS00008 92->98 MYRISTYL PDOC00008

(No Pfam data available for DKFZphtes3\_2f14.2)



DKF2phtes3\_2g7

group: testes derived

DKF2phtes3\_2g7 encodes a novel 359 amino acid protein with similarity to neurofilament proteins.

No informative BLAST results; No predictive prosite, pfam or SCOP motive.

The new protein can find application in studying the expression profile of testis-specific genes.

similarity to neurofilament proteins

complete cDNA, complete cds, 6 EST hits (5 hits are out of a testis library)

Sequenced by EMBL

Locus: unknown

Insert length: 1613 bp

Poly A stretch at pos. 1595, polyadenylation signal at pos. 1557

```
1 GCCACACAGG CTCCTTGGAG TAAGAGTGTG AGAAACTGGA TGAAGACAGC
51 TGTATTCTTT TGAAGCGTTC CGAGATTGGT CTGTCTCTAC CAACTAAAAA
101 CTTCTAGCTT AAGTGCAGAG ATTTAAGGAG ATCAACAAAA ACTCAGTCTA
151 GACATATTAT GAGGCTGGGA GGGTATCAAC AGACTTGAGT TCTTGTGACG
201 AAGATCACCT GCTTTTAATA TTGTCTCAG GGTCTGAGCA CATCTGGAAG
251 TGAGGTCAAT CAAGTTAGAC CCCAAAAACT TTTGTGACAA CAGTGAAGAG
301 GGGAAAAATA ACACACCACA AACATGAACC TCAACCCCCC GACATCTGCT
351 CTTCAGATCG AGGGCAAAGG CAGCCATATT ATGGCTAGAA ATGTAAGCTG
401 CTTCTAGTGC AGGCACACCC CTCATCCCAG AAGAGTCTGC CACATCAAAG
451 GCTTGAATAA CATTCCAATC TGTACTGTGA ATGATGATGA GAATGCATTT
501 GGAACATTGT GGAAGTTGG CCAGTCTAAC TACTTAGAGA AGAACAGGAT
551 ACCATTTGCC AATTGCAGTT ACCCCCCGAG CACTGCAGTC CAGAAGAGCC
601 CTGTAAGAGG AATGTCGCCA GCCCCAAACG GTGCCAAAGT GCCTCCACGG
651 CCTCATCTCG AGCCCACTAG AAAAATTAAA GAGTGTCTCA AAACCTCCAG
701 TGAGAATCCC TTAGTAATTA AAAAGGAAGA AATTAAGGCC AAAAGACCAC
751 CATCACCTCC AAAGGCATGC TCTACTCCTG GCTCCTGTTC TTCAGGGATG
801 ACAAGTACCA AGAATGATGT GAAAGCAAAC ACCATTGTGA TACCAAACTA
851 TCTGGATCAG GAAATAAAAA TCCTGGCAAA GCTCTGTAGC ATTTTGCATA
901 CTGATTCTCT GGCAGAAGTT TTACAGTGGC TGCTTCATGC AACTTCAAAA
951 GAAAAAGAGT GGGTCTCAGC TTTGATTCAT TCTGAGCTTG CCGAGATAAA
1001 CCTGTAACT CATCACAGAA GAAACACCTC AATGGAACCA GCAGCAGAGA
1051 CTGGGAAGCC ACCCACAGTT AAATCACCAC CCACAGTTAA ATTGCCCCCA
1101 AATTTTACTG CAAAATCAAA AGTGCTGACC AGAGATACAG AAGGGGATCA
1151 ACCAACCAGA GTGTCAAGTC AAGGATCTGA AGAAACAAAG GAAGTACCAA
1201 AAGAGGCTGA GCACAAGCCT CCACTACTTA TAAGAAGAAA TAATATGAAA
1251 ATACCTGTTG CAGAATATT CAGCAACCA AATTCTCCTC CCAGGCCTAA
1301 CACTCAGGAG AGTGGATCAG CAAAACCACT GTCAGCAAGG AGTATACAAG
1351 AATACAACCT CTGTCCCCAA AGAGCATGTT ATCCTTCAAC ACACCGGAGG
1401 TAGAAGTTCT AGACTGGGTG AATTCTTTCA TGAATATGAG CTTACATTTT
1451 ACATCATCAA ATTATTTTTC AAATGAATAT TTTTGTATT GAGGAATCAA
1501 GTGGTCTCTT TTATGGTGGC ACATGTAAT CTAAAAATAC CTGTATGTAA
1551 TGCTACAAAT AAATATTACT GGAATGATA TTTCCATTG TAGTTAAAAA
1601 AAAAAAAAAA AAA
```

## BLAST Results

No BLAST result

## Medline entries

No Medline entry

## Peptide information for frame 3

ORF from 324 bp to 1400 bp; peptide length: 359  
Category: similarity to known protein

```

1 MNLNPPTSAL QIEGKGSIM ARNVSCFLVR HTPHPRRVCH IKGLNNIPIC
51 TVNDDENAFG TLWEVQGSNY LEKNRIPFAN CSYPPSTAVQ KSPVRGMSPA
101 PNGAKVPPRP HSEPSRKIKE CFKTSSSENPL VIKKEIKAK RPPSPPKACS
151 TPGSCSSGMT STKNDVKANT ICIPNYLDQE IKILAKLCSI LHTDSLAEVL
201 QWLLHATSKE KEWVSALIHSE LAEINLLTH HRRNTSMEPA AETGKPPTVK
251 SPPTVKLPPN FTAQSKVLTR DTEGDQPTRV SSQGSSEENKE VPKEAEHKPP
301 LLIRNNMKI PVAEYFSKPN SPPRPNTQES GSAKPVARS IQEYNLCPPQ
351 ACYPSTHRR

```

## BLASTP hits

Entry A43427 from database PIR:  
 neurofilament triplet H1 protein - rabbit (fragment)  
 Score = 118, P = 5.6e-04, identities = 79/290, positives = 110/290

Entry RNNFH\_1 from database TREMBL:  
 Rat high molecular weight neurofilament (NF-H) protein mRNA, 3' end.  
 Score = 115, P = 9.5e-04, identities = 69/281, positives = 100/281

Entry B43427 from database PIR:  
 neurofilament protein H form H2 (repetitive region) - rabbit (fragment)  
 Score = 111, P = 1.3e-03, identities = 64/269, positives = 102/269

Alert BLASTP hits for DKFZphtes3\_2g7, frame 3

No Alert BLASTP hits found

Pedant information for DKFZphtes3\_2g7, frame 3

## Report for DKFZphtes3\_2g7.3

```

[LENGTH]      359
[MW]           39725.53
[PI]           9.45
[PROSITE]      MYRISTYL      3
[PROSITE]      CAMP_PHOSPHO_SITE      1
[PROSITE]      CK2_PHOSPHO_SITE      9
[PROSITE]      PKC_PHOSPHO_SITE     10
[PROSITE]      ASN_GLYCOSYLATION      4
[KW]           Alpha_Beta
[KW]           LOW_COMPLEXITY      4.18 %

```

```

SEQ  MNLNPPTSALQIEGKGSIMARNVSCFLVRHTPHPRRVCHIKGLNNIPICTVNDDENAFG
SEG  .....
PRD  ccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc

SEQ  TLWEVQGSNYLEKNRIPFANCSYPPSTAVQKSPVRGMSAPNGAKVPPRPHSEPSRKIKE
SEG  .....
PRD  cccccccccccccccccccccccccccccccccccccccccccccccccccccccccchhhhhh

SEQ  CFKTSSSENPLVIKKEIKAKRPPSPPKACSTPGSCSSGMTSTKNDVKANTICIPNYLDQE
SEG  .....
PRD  hccccccccccccccccccccccccccccccccccccccccccccccccccccccccchhh

SEQ  IKILAKLCSILHTDSLAEVLQWLLHATSKEKEWVSALIHSELAEINLLTHHRRNTSMEPA
SEG  .....
PRD  hhhhhhhhhhhccccchhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhcccccccc

SEQ  AETGKPPTVKSPPTVKLPPNFTAKSKVLTRDTEGDQPTRVSSQGSSEENKEVPKEAEHKPP
SEG  .....
PRD  ccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc

SEQ  LLIRNNMKIPVAEYFSKPNPPRPNTQESGSAKPVARSIQEYNLCPPQACYPSTHRR
SEG  .....
PRD  eeeeeccccccccccccccccccccccccccccccccchhhhhcccccccccccccccccc

```

## Prosites for DKFZphtes3\_2g7.3

```

PS00001      23->27      ASN_GLYCOSYLATION      PDOC00001
PS00001      80->84      ASN_GLYCOSYLATION      PDOC00001
PS00001      234->238    ASN_GLYCOSYLATION      PDOC00001

```

WO 01/12659

PCT/IB00/01496

PS00001	260->264	ASN_GLYCOSYLATION	PDOC00001
PS00004	232->236	CAMP_PHOSPHO_SITE	PDOC00004
PS00005	115->118	PKC_PHOSPHO_SITE	PDOC00005
PS00005	161->164	PKC_PHOSPHO_SITE	PDOC00005
PS00005	207->210	PKC_PHOSPHO_SITE	PDOC00005
PS00005	243->246	PKC_PHOSPHO_SITE	PDOC00005
PS00005	248->251	PKC_PHOSPHO_SITE	PDOC00005
PS00005	254->257	PKC_PHOSPHO_SITE	PDOC00005
PS00005	262->265	PKC_PHOSPHO_SITE	PDOC00005
PS00005	332->335	PKC_PHOSPHO_SITE	PDOC00005
PS00005	337->340	PKC_PHOSPHO_SITE	PDOC00005
PS00005	356->359	PKC_PHOSPHO_SITE	PDOC00005
PS00006	51->55	CK2_PHOSPHO_SITE	PDOC00006
PS00006	61->65	CK2_PHOSPHO_SITE	PDOC00006
PS00006	124->128	CK2_PHOSPHO_SITE	PDOC00006
PS00006	162->166	CK2_PHOSPHO_SITE	PDOC00006
PS00006	195->199	CK2_PHOSPHO_SITE	PDOC00006
PS00006	207->211	CK2_PHOSPHO_SITE	PDOC00006
PS00006	235->239	CK2_PHOSPHO_SITE	PDOC00006
PS00006	272->276	CK2_PHOSPHO_SITE	PDOC00006
PS00006	340->344	CK2_PHOSPHO_SITE	PDOC00006
PS00008	153->159	MYRISTYL	PDOC00008
PS00008	158->164	MYRISTYL	PDOC00008
PS00008	284->290	MYRISTYL	PDOC00008

(No Pfam data available for DKFzphtes3\_2g7.3)

DKFZphtes3\_2h1  
-----

group: transmembrane protein

DKFZphtes3\_2h1 encodes a novel 116 amino acid protein with weak similarity to C. elegans cosmid C13F10.

The novel protein contains 1 transmembrane region.  
No informative BLAST results; No predictive prosite, pfam or SCOP motive.

The new protein can find application in studying the expression profile of testis-specific genes and as a new marker for testicular cells.

similarity to C.elegans C13F10.5

TRANSMEMBRANE 1

Sequenced by EMBL

Locus: /map="2"

Insert length: 1156 bp

Poly A stretch at pos. 1143, polyadenylation signal at pos. 1121

```

1  GGCCATCAAA ATAACATAAC CATGTCATTT GGAGCAACAA AGCCACTGCG
51  GGCTCCATTT GGGCCAAGCT CTGACTGCAA TGATGCCTCT GCCCGACCC
101  GGGCCTCGCT GTGACTGACA ATGCCGCTGC ATCTTTTCAG CAGTCATTGA
151  TGAGGAAGTA TCTACATCCT CCTTCCCACT ACCAGATTTT GCTTGGAGAA
201  AAGCAGTTTC CTGAAATAAT TCTGTGACGA GCTTCTTCCA CATTAGGACA
251  AAAATGCTGG AAGCGGCTCA GCCCAGGGC AGCACATCAG AGACACCATG
301  GAACACAGCC ATTCCTCTGC CGTCGTGCTG GGACCAGTCT TTCCTGACCA
351  ATATCACCTT CTGGAAGGTT CTCTCTGGT TGGTCCTGCT GGGACTGTTT
401  GTGGAACGTT AATTGGCCTT GGCATATTTT GTCCTGCTCT TGTTCATTG
451  GATGTACGTC GGGACACGAG GCCCTGAAGA GAAGAAAGAG GGAGAGAAGA
501  GCGCCTACTC TGTGTTCAAT CCAGGCTGTG AAGCCATCCA GGGCACCCCTG
551  ACTGCAGAGC AGTTGGAGCG CGAGTTACAG TTGAGACCCC TGGCAGGGAG
601  ATAGGACCCA GCTGTGCTGT CATGCAGCTA ACCTCTGATG TGGTCTTCCT
651  CACCATTTGG TATGGATTG ATTCAGGTG TATAGGACTA AGGGCAGCTT
701  GCGGGTTAGC TCTGTGACTG CATAGTTTTT CTACCTTCTT TCCCTGATCT
751  TTTGCTGCCA TTTGATCTTT GATAGTTTTG GTGAAACTCT CTAATAATCA
801  TTCCTGTGG GTCCGACGCA ATTTATAAAA ATTATGTACT CAAGAAGGGA
851  GACCTGTTTG TTTCAATCTT CATCTGTTTG GGAGATGATT TTAGAGCACT
901  AGAAAGGCAC TGGGGAGATT CTCAGCTTAA AACATCCAGC AGTTTGAAGT
951  ATGATTAGGT ACATCAGGGC TGCATTGTCA ATGTCTCTTT TAAGTCTTTT
1001 AACATTTATA GCAATTTTTT TTTTCCCGGA GAGTTTAGGT TGCAAGTTTT
1051 GGGTTTCTTG TTTGTTTTTG TTTTGCTTCC TGCTTTAATT CTTTAATTTT
1101 CAGTCATTAC TGGTATTGAA AAATAAAATA TCTTAAAC ATCAAAAAAA
1151 AAAAAA

```

#### BLAST Results

-----

Entry HS313307 from database EMBL:

human STS SHGC-16715.

Score = 1222, P = 1.4e-48, identities = 248/251

#### Medline entries

-----

No Medline entry

#### Peptide information for frame 2

-----

ORF from 254 bp to 601 bp; peptide length: 116

Category: similarity to unknown protein

```

1  MLEAAQPGS TSETPWNTAI PLPSCWDQSF LTNITFLKVL LWLVLLGLFV
51  ELEFGLAYFV LSLFYWMYVG TRGPEEKKEG EKSAYSVFNP GCEAIQGLT
101  AEQLERELQL RPLAGR

```

No BLASTP hits available

TREMBL:CEUC13F10.2 gene: "C13F10.5"; *Caenorhabditis elegans* cosmid C13F10., N = 1, Score = 141, P = 8.2e-10

**HSPs:**

Query: 87 VFNPGCEAIQGTLTAEQLEREL 108  
VFN CE + G++TAE ER++  
Sbjct: 145 VFNDNCERLAGSMTAEHFERDM 166

Report for DKFZphtes3\_2h1.2

```

SEQ      MLEAAQPGQSTSETFPWNTAIPLPSCDQDSFLTNTIFLKVLVLVLLGLFVELEFGLAYFV
PRD      .....xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx
PRD      cccccccccccccccccccccccccccccccccccccccccchhhhhhhhhhhhhhhhhhhhhhhhhhh
MEM      .....MMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMM.....

SEQ      LSLFYWMYVGTGRGPEEKKEGEKSAYSVFNPGECAIQTGLTAEQLERELQLRPLAGR
PRD      .....xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx
PRD      hhhhhhhhhcccccchhhhhhhcccccceccccccecccccchhhhhhhhhhhhhccccc
MEM      .....

```

PS00001	33->37	ASN_GLYCOSYLATION	PDOC00001
PS00006	10->14	CK2_PHOSPHO_SITE	PDOC00006
PS00006	24->28	CK2_PHOSPHO_SITE	PDOC00006
PS00007	78->86	TYR_PHOSPHO_SITE	PDOC00007
PS00007	77->86	TYR_PHOSPHO_SITE	PDOC00007
PS00008	97->103	MYRISTYL	PDOC00008

797

DKF2phtes3\_2h15  
-----

group: testes derived

DKF2phtes3\_2h15 encodes a novel 855 amino acid protein with very weak similarity to *S. pombe* cdc23.

No informative BLAST results; no predictive prosite, pfam or SCOP motif.

The new protein can find application in studying the expression profile of testis-specific genes.

similarity to cdc23

complete cDNA, complete cds, EST hits

Sequenced by EMBL

Locus: unknown

Insert length: 4619 bp

Poly A stretch at pos. 4598, polyadenylation signal at pos. 4589

```
1 GAAGCGCTCC CGGCATCGGC CAAGATTCTA CATTGCTCAT CTGGGCATCT
51 GAGCCTCCTT CGAAGTTTCC TGTCACAACT GTCCTCTTGA CAGCATGGAT
101 GAGGAGGAAG ACAATCTGTC TCTGCTGACC GCACTGCTGG AAGAAAATGA
151 GTCAGCCTTG GATTGTAATT CAGAAGAAAA TAACCTCTTG ACGCCGGAAA
201 ATGGCGAGCC CGACGCATTG GATGAGCTCT TTGATGCCGA CGGCCAGCGT
251 GAATCTTATA CAGAAGAGGC TGATGATGGA GAAACAGGAG AGACAAGAGA
301 CGAAAAGGAA AATCTGGCCA CTCTCTTTGG AGATATGGAG GACTTAACAG
351 ATGAAGAGA AGTTCGCCCA TCACAGTCAA CTGAAAATAG GGTCTCCCT
401 GCTCCTGCCC CCAGGCGAGA GAAAACGAAT GAAGAGTTGC AAGAGGAATT
451 AAGGAATTTG CAAGAGCAAA TGAAGGCTT ACAAGAGCAG CTAAGAGTAA
501 CAACAATTAA ACAGACAGCA AGCCCAGCCC GTCTGCAAAA ATCCCCTGAG
551 AAGTCTCCCC GGCCACCTCT TAAGGAGAGG AGAGTTCAGA GAATTCAGGA
601 GTCAACATGC TTTTCTGCGG AGCTTGATGT CCTGCGCTA CCAAGAACCA
651 AGAGGGTGGC TCGAACACCA AAGCCTTCAC CTCCAGATCC CAAAAGCTCA
701 TCTTCAAGGA TGACAAGTGC ACCCTCCCAA CCCCTACAGA CGATTTCTCG
751 GAACAAACCT AGTGGGATAA CTAGAGGTCA AATTGTGGGG ACCCCAGGAA
801 GTTCTGGGGA AACGACTCAA CCCATCTGTG TGGAAAGCTT CTCTGGTCTG
851 CGGCTCAGGC GGCCTCGAGT ATCCTCCACA GAAATGAACA AGAAAATGAC
901 CGGCCGAAAA CTGATCAGAC TGTCTCAGAT CAAGGAAAAA ATGGCCAGAG
951 AGAAGCTGGA AGAATAGAT TGGGTGACAT TTGGGGTTAT ATTGAAGAAG
1001 GTTACGCCAC AGAGTGTGAA TAGTGGAAAA ACCTTCAGCA TATGGAAGCT
1051 GAATGATCTT CGTGACCTGA CACAATGTGT GTCCTTGTTT TTATTGGAG
1101 AAGTTCACAA AGCGCTCTGG AAGACGGAGC AGGGGACTGT CGTAGGGATC
1151 CTCATAGCCA ACCCATGAA GCCCAAGGAT GGTTCAGAGG AGGTGTGTTT
1201 ATCTATCGAT CATCTCAGA AGGTCTTAAT TATGGGTGAA GCTCTTGACC
1251 TGGGAACCTG TAAAGCCAAG AAGAAGATG GAGAGCCGTG CACGCAGACT
1301 GTGAATTTGC GTGACTGTGA GTACTGTGAG TACCATGTCC AGGCTCAGTA
1351 CAAGAAGCTC AGTGCAAAGC GTGCGGATCT GCAGTCCACC TTCTCTGGAG
1401 GACGAATTCC AAAGAAGTTT GCCCGCAGAG GCACCAGCCT CAAAGAACGG
1451 CTGTGCCAAG ATGGCTTTTA CTACGGAGGG GTTCTTCTG CCTCGTATGC
1501 AGCTTCAATT GCAGCAGCTG TGGCTCCTAA GAAGAAGATT CAAACCATC
1551 TGAGTAATCT GGTGTGTAAG GGCACAAACT TGATCATCCA GGAACACAGG
1601 CAAAACCTCG GAATACCCCA GAAGAGCCTG TCTTGCTCTG AGGAGTTCAA
1651 GCAACTGATG GACCTGCCGA CGTGTGGAGC CAGGAACCTA AAACAACATT
1701 TAGCCAAAGC CTCAGCTTCA GGGATTATGG GGAGCCCAA ACCAGCCATC
1751 AAGTCCATCT CGGCCTCAGC ACTCTTGAAG CAACAGAAGC AGCGGATGTT
1801 GGAGATGAGG AGAAGGAAAT CAGAAGAAAT ACAGAAGCGA TTTCTGCAGA
1851 GCTCAAGTGA AGTTGAGAGC CCAGCTGTGC CATCTTCATC AAGACAGCCC
1901 CCTGCTCAGC CTCCACGGAC AGGATCCGAG TTCCCCAGGC TGGAGGGAGC
1951 CCCGGCCACA ATGACGCCCA AGCTGGGGCG AGGTGTCTTG GAAGGAGATG
2001 ATGTTCTCTT TTATGATGAG TCACCACCAC CAAGACCAA ACTGAGTGCT
2051 TTAGCAGAAG CCAAAAAGTT AGCTGCTATC ACCAATTAA GGGCAAAAGG
2101 CCAGGTTCTT ACAAAAACAA ACCCAAACAG CATTAAAGAG AAACAAAAGG
2151 ACCCTCAGGA CATCTGGAG GTGAAGGAC GTGTAGAAAA AAACACCATG
2201 TTTTCTCTC AAGCTGAGGA TGAATTGGAG CCTGCCAGGA AAAAAAGGAG
2251 AGAACAACTT GCCTATCTGG AATCTGAGGA ATTTAGAAA ATCTTAAAG
2301 CAAATCAAAA ACACACAGGC ATCTGAAAG AGGCCGAGGC TGAGATGCAG
2351 GAGCGCTACT TTGAGCCACT GGTGAAAAAA GAACAAATGG AAGAAAAGAT
2401 GAGAAACATC AGAGAAGTGA AGTGCCGTGT CGTGACATGC AAGACGTGCG
2451 CCTATACCCA CTTCAGCTG CTGGAGACCT GCGTCAGTGA GCAGCATGAA
2501 TACCACCTGGC ATGATGGTGT GAAGAGGTTT TTCAATATGC CTTGTGAAA
2551 CAGAAGCATC TCCTTGGACA GACTCCCGAA CAAGCACTGC AGTAACGTGT
2601 GCCTCTACAA ATGGGAACGG GACGGAATGC TAAAGGTATG CCATTTGCGT
2651 ACTAATTTTT GACTCCTTTT AGTGACCCAT GCTAATAATG TGAACCATC
```

```

2701 TCCTATTAAA ATATTTTCAT TTTTCTAGGA AAAGACTGGT CCAAAGATAG
2751 GAGGAGAAAC TCTGTTACCA AGAGGAGAAG AACATGCTAA ATTTCTGAAC
2801 AGCCTTAAAT AACCCGAACT TCAGACATTT TCCCACAGAC TTCCTGGCCT
2851 CCTGTGACTC TGGAAAGCAA AGGATTGGCT GTGTATTGTC CATTGATTCC
2901 TGATTGACGC CGTCAAAAAC AAATGCTTGT TAAGCCCATA AGCTTTGCCCT
2951 GCTTACTTTC TGCCATTGGG TTGGTTTGAT ACCACATTTA ACATTGACAT
3001 TTAAGTGGAA AACCAAGTTA TCATTGTCTT TTCTAAGCTC AGTGTGGATG
3051 ATTGCATTAC TTCATTCACT GAAGTTTTTG CCAAAAAATT GGAAGGTAAA
3101 CAGAGAGCTA TGTTTCTGTA TCTTTTGGTT ATAGAGTGTT CACTTCTTTA
3151 TCATAACAAA ATTCTAGTGT TTATACGAAC ACCCAGAGGC AAAAGAATTT
3201 GGCTTAATTC TCACTCCAGG TAAGTAGCTT AACTTCTGGG CTTCAGTTTT
3251 CTCATCTGTA AAATCAGGAA GATTGGACTA AGTGATCCTG AAATGTATTT
3301 TTTAGCACTG GATTCTACA AATAATAAAA CTTTCCCATC TAGATAATGA
3351 TGATCACATA GTCTTGATGT ACGGACATTA AAAGCCAGAT TTCTTCATTC
3401 AATTCTGTGA TCTCTGTTTT ACTCTTTGAA ATTGATCAAG CCAGTGAATC
3451 ACTTTGCATT TCAGTTTATA TATAGAGAGA GAAAGAAGGC TGTCTGCTCT
3501 TACATTATTG TGGAGCCCTG TGATAGAAAT ATGTAAAATC TCATATTATT
3551 TTTTTTTTAA TTTTTTTATT TTTTATGACA GGGTCTCACT ATGTCACCCT
3601 GGCTGGAGTG CAGTAGTGCG ATCGCGGCAC ACTGCAGCCT TGGCTTCCCT
3651 GGGCTCAAGC AGTCCTCCCA CCTCAGTCTC CCAATAGCT AGGACTACAG
3701 GCGTGCCTGA CCAAGCCAG CTAATTTTTG CATTTTTTGT AGAGATGGGG
3751 TTTTGCCATG TTGCTCAGGC TGGTCTCAAA CTCCTGAGCA CTAGCAATCC
3801 ACCCACCTCT GTTCCAAAA AAAAAAAAAA AATGAAAGGT CAACCCCTAT
3851 GCAATATTACC ACAGCAAAGG TTTCATTGAG GAGATTCTTC CATCTGGGCA
3901 ACCTGGTTTT CCAATATCA TTTGACCTAA GTGAATGTTG ATACTAGCTA
3951 AAGATTGGGT AAATTGGTTG AATTATTGTA TTGAAGCTTG AGCTGTAGCT
4001 AAAAGTAATT TAGGTTTCCC CTAAGATGTT ATTATGTTAG GGACATAACA
4051 CTTTGGGAG GTTGTGTGGG GAGATGGTTG ATTTAGGTTT TCAAAAGCTA
4101 GAAATAAAAT TTACATGCCT TAGATTTCAT AAAATTCTGC TCTAATTGGG
4151 TGAAGGTGTC TGTATCTAAC TTGTGTTCCCT CCTAAGGTTA TGTCTTAATA
4201 ACTATTCTTT TAGGAGTATA CTTCTACTTT ATAGAAGGTT GCTTTTCTTT
4251 TTAATTTTTT CTAACAAAGA AAAGAATAAA GTATTATTAT ATAAGAACCA
4301 GAAAGCACTT GAAACTGATG TTTTAAATGG CTCATTTAGG GTAGATTAT
4351 TTATCTCATT AACTTAAAC AGCTATGTGT ATGAAATAGG TCACAACAGA
4401 ACTTGAACAC CAGGTTGGTG TCTGAGCAAT CCCTTTCTTA TGGGAAAAAC
4451 AATGTTCTTG TTTGAACAGA GGGTATCATT GCAGTCAGTA TTCACGTGTA
4501 TATTGTTATA TAAGTTGTAT AATATGCTTG TAAAGGCTGA GGGTGAGCTG
4551 TATCTGGATG CCTTTTACA ATTTGATTTT AACTTTTAAA ATAAATTTAA
4601 AACATAAAAA AAAAAAAAAA

```

## BLAST Results

No BLAST result

## Medline entries

No Medline entry

## Peptide information for frame 2

ORF from 95 bp to 2659 bp; peptide length: 855  
 Category: similarity to known protein  
 Classification: Cell division

```

1 MDEEEDNLSL LTALLEENES ALDCNSEENN FLTRENGEPD AFDELFDADG
51 DGESYTEEAD DGETGETRDE KENLATLFGD MEDLTDEEEV PASQSTENRV
101 LPAPAPRREK TNEELQEELR NLQEQMKALQ EQLKVTTIKQ TASPRLQKS
151 PEKSPRPPLK ERRVQRIQES TCFSAEILDVP ALPRTKRVAR TPKPSPPDPK
201 SSSSRMTSAP SQPLQITSRN KPSGITRGQI VGTPGSSGET TQPICVEAFS
251 GLRLRRPRVS STEMNKKMTG RKLIRLSQIK EKMAREKLEE IDWVTFGVIL
301 KKVTPQSVNS GKTFSIWKLN DLRDLTQCVS LFLFGEVHKA LWKTEQGTUV
351 GILNANPMKP KDGESEEVCLS IDHPQKVLIM GEALDLGTCK AKKNGEPTV
401 QTVNLRDCEY CQYHVQAQYK KLSAKRADLQ STFSGGRIK KFAARRGTSK
451 ERLCQDGFY GGVSSASYAA SIAAAVAPKK KIQTLSNLV VKGTNLIQE
501 TRQKLGIPOK SLSCSEEFKE LMDLPTCGAR NLKQHLAKAS ASGIMGSPKP
551 AIKSISASAL LKQKQRMLE MRRRKSEIQQ KRFLQSSSEV ESPAVPSSSR
601 QPPAQPPRTG SEFPRLGAP ATMTPKLGRG VLEGDDVLFY DESPPRPKL
651 SALAEAKKLA AITKLRAKGQ VLTKTNPSNI KKKQKDPQDI LEVKERVEKN
701 TMFSSQAEDE LEPARKKRE QLAYLESEEF QKILKAKSKH TGILKEAEAE
751 MQERYFEPLV KKEQMEEKMR NIREVKCRV TCKTCAYTHF KLETCVSEQ
801 HEYHWDGVK RFFKPCGGR SISLDRLPNK HCSNCGLYKW ERDGMKLVCH
851 LRTNF

```

## BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphtes3\_2h15, frame 2

TREMBLNEW:SPBC1347\_10 gene: "cdc23"; "SPBC1347.10"; product: "cell division cycle protein 23"; S.pombe chromosome II cosmid c1347., N = 2, Score = 284, P = 7e-21

PIR:S48384 DNA43 protein - yeast (Saccharomyces cerevisiae), N = 2, Score = 203, P = 7e-12

TREMBL:SCDNA52A\_1 gene: "DNA52"; Saccharomyces cerevisiae DNA52 gene, complete cds., N = 2, Score = 201, P = 7.9e-12

TREMBLNEW:AC006234\_6 gene: "F5H14.6"; Arabidopsis thaliana chromosome II BAC F5H14 genomic sequence, complete sequence., N = 2, Score = 211, P = 1.7e-15

PIR:S48384 DNA43 protein - yeast (Saccharomyces cerevisiae), N = 2, Score = 203, P = 7.2e-12

>TREMBLNEW:SPBC1347\_10 gene: "cdc23"; "SPBC1347.10"; product: "cell division cycle protein 23"; S.pombe chromosome II cosmid c1347.  
Length = 593

## HSPs:

Score = 284 (42.6 bits), Expect = 7.0e-21, Sum P(2) = 7.0e-21  
Identities = 97/383 (25%), Positives = 186/383 (48%)

Query: 109 EKTNEELQEELRNLEQEMKALQEQLKVTTIKQTASPARLQKSPKSPRPPLKERRVQRIQ 168  
E+ + +L+E + LQ Q+ +QE+ ++ + + AS + + PR P ++ RV +  
Sbjct: 8 EENDLDLEE--KRLQRLNEIQEKKRLRSAQKEASSENAEVI--QVPRSPQQVRVLTVS 63

Query: 169 ESTCFSAE----LDVPALPRTKRVARTPKPSPDPKSSSSRMTSAPSQP-----LQTIS 218  
+ + + L + + K V+ P P PK R+ A +Q L+T+  
Sbjct: 64 SPSKLSKPKRLILGIDKGTGKDVSLGKGRGPLPKPFHERLAEARNQERKRSCLKTKMK 123

Query: 219 RNKPSGITRGQIVGTPGSSGETTQPI-C--VEAFSGRLRRPRVSSTEMNKKMTGRKLIR 275  
+N+ R + + G S E P+ C ++ +S + +S + + G ++  
Sbjct: 124 KNRKQSFQRKRNILEDGKSEEEKFPMKCDEIDPYSRQAIVIRYISDEVAKENIGGNQVYL 183

Query: 276 LSQIKEKMAREKLE--EID-WVTFGVILKKV-TPQSVNSGKTFSIWKLNDLRDLTQCVSL 331  
+ Q+ + + K E E+D +V G++ T ++VN K + + L DL+ +C  
Sbjct: 184 IHQLLLVRAPKFEAPEVDNYVVMGIVASNSGTRETVNGNK-YCMLTLTDLKWQLEC--- 239

Query: 332 FLFGEVHKALWKEQGTVVGILNANPMKPKDGS--EEVCLSIDHPQKVLII-MGEALDLGTC 389  
FLFG+ + WK + GTV+ +LN +KPK+ L +D VL+ +G + LG C  
Sbjct: 240 FLFGKAFERYWKIQSGTVIALLNPEVLKPKNPDIGRFSCLKDSEYDVLLEIGRSKHLGYC 299

Query: 390 KAKKKNGEPTQTQTNLRDCEYCOYHVOAQYKKLSAKRADLQSTFSGGRIPKKFARRGTS 449  
+++K+GE C ++ R + C+YHV ++ + R + S+ + P+ ARR  
Sbjct: 300 SSRKSGELCKHWDKRGADVCEYHVDLAVQRSMSTRTEFASSMATMHEPR--ARR---- 353

Query: 450 KERLCQDGF--YYGGVSSASYAASIAAAVAPKKKIQT 484  
++R GF Y+ G ++ +A + +QT  
Sbjct: 354 EKRRFGQGFQGYFAGEKYSAIPNAVAGLYDAEDAVQT 390

Score = 41 (6.2 bits), Expect = 7.0e-21, Sum P(2) = 7.0e-21  
Identities = 12/43 (27%), Positives = 17/43 (39%)

Query: 453 LCQDGFYYGGVSSASYAASIAAAVAPKKKIQTTLNLVVKGTN 495  
L +D S AS A++ K + SN + GTN  
Sbjct: 465 LSKDSEIDSSTKKPSVLASFNASIMNPKSSLPSFSNSAILGTN 507

Score = 40 (6.0 bits), Expect = 8.9e-21, Sum P(2) = 8.9e-21  
Identities = 13/26 (50%), Positives = 18/26 (69%)

Query: 536 LAKASASGIMSGPKPAIKSISASALL 561  
LA +AS IM +PK ++ S S SA+L  
Sbjct: 481 LASFNAS-IM-NPKSSLPSFSNSAIL 504

Pedant information for DKFZphtes3\_2h15, frame 2

Report for DKFZphtes3\_2h15.2



[LENGTH] 855  
[MW] 96135.01  
[PI] 8.96  
[HOMOL] TREMBLNEW:SPBC1347\_10 gene: "cdc23"; "SPBC1347.10"; product: "cell division  
cycle protein 23"; S.pombe chromosome II cosmid c1347. 5e-16  
[FUNGAT] 03.22 cell cycle control and mitosis [S. cerevisiae, YIL150c] 1e-11  
[FUNGAT] 03.16 dna synthesis and replication [S. cerevisiae, YIL150c] 1e-11  
[FUNGAT] 30.10 nuclear organization [S. cerevisiae, YIL150c] 1e-11  
[KW] Alpha\_Beta  
[KW] LOW\_COMPLEXITY 12.05 %  
[KW] COILED\_COIL 4.21 %

SEQ MDEEDNLSLLTALLEENESALDCNSEENNFLTRENGEPDAFDELFDADGDGESYTEEAD  
SEG .....xxxxx  
PRD cccchhhhhhhhhhhhhhhhhcc  
COILS .....

SEQ DGETGETRDEKENLATLFGDMEDLTDEEEVPASQSTENRVLPAPAPRREKTNEELQEELR  
SEG .....xxxxxxx  
PRD cccccccccchhhhhhhcc  
COILS .....CCCCCCCCCCCC

SEQ NLQEQMKALQEQLKVTTIKQTASPARLQKSPEKSPRPPLKERRVQRIQUESTCFSAELDVP  
SEG .....xxxxx  
PRD hhhhhhhhhhhhhhhhhhhhhcc  
COILS CCCCCCCCCCCCCCCCCCCCCC.....

SEQ ALPRTKRVARTPKPSPDPKSSSRMTSAPSQPLQTSIRNKPSGITRGQIVGTPGSSGET  
SEG .....xxxxxxx  
PRD cccccccccccccccccccccchhhhhhhccccchhhhhcccccccccccccccccccccccc  
COILS .....

SEQ TQPICVEAFSGLRLRRPRVSSTEMNKKMTGRKLIRLSQIKEKMAKLEIDWVTFGVIL  
SEG .....  
PRD cccccccccchhhhhhhccccchhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhccceeeeeee  
COILS .....

SEQ KKVTPQSVNSGKTFISIWLNDLRLDTCQVSLFLFGEVHKALWKEQTGVVILNANPMKP  
SEG .....  
PRD cccccccccccccccccccccchhhhhhhheeeecchhhhhhhccceeeeecccccccc  
COILS .....

SEQ KDGSEEVCLSIDHPQKVLIMGEALDLGTCAKKKNGEPTQTVNLRDCEYCYHVQAQYK  
SEG .....  
PRD cchhhhhhhhh  
COILS .....

SEQ KLSAKRADLQSTFSGGRIPKKFARRGTSLSKERLCQDGFYGGVSSASAYASIAAAVAPKK  
SEG .....xxxxxxx  
PRD hhhhhhhhhhhccccccccccccccccccccchhhhhhhccccccccccccchhhhhhhhhhhhhcch  
COILS .....

SEQ KIQTTLNVLVVGKTNLIQETROQLGIPQKSLSCSEEFKELMDLPTCGARNLKQHLAKAS  
SEG .....  
PRD hhhhhhhheccccccccchhhhhhhccccccccchhhhhhhhhccccccccchhhhhhhhhhh  
COILS .....

SEQ ASGIMGSPKPAIKSISASALLKQKQRMLEMRRRKSEEIQKRFLQSSSEVESPAVPSSSR  
SEG .....xxxxxxx  
PRD hhccccccccccccchhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhcccccccccccc  
COILS .....

SEQ QPPAQPPRTGSEFFRLEGAPATMTPKLGRLVLEGDDVLFYDESPPPRPKLSALAEAKKLA  
SEG .....xxxxxxx  
PRD cchhhhhhhhhhhhh  
COILS .....

SEQ AITKLRAKGVLTNTNPNSIKKKQKDPQDILEVKEKVEKNTMFSSQAEDELEPARKKRRE  
SEG .....xxxxx  
PRD hhhhhhhheeeccccccccccccccccchhhhhhhhhhhhhhhccchhhhhhhhhhhhhhhhh  
COILS .....

SEQ QLAYLESEEFQKILKAKSKHTGILKEAEAMQERYFEPLVKKEQMEKMRNIREVKCRVV  
SEG .....  
PRD hhhhhhhhhhhhhhhccccchheee  
COILS .....

SEQ TCKTCAYTHFKLLETVCVSEQHEYHWDGVKRFKPCGNRSISLDRLPNKHCSNCGLYKW

WO 01/12659

PCT/IB00/01496

SEG .....  
PRD eeecceeeeeeccccceeeccccccccceeeccccccccccccccccccccceec  
COILS .....

SEQ ERDGMLKVCHLRTNF  
SEG .....  
PRD ccccccccccccccc  
COILS .....

(No Prosite data available for DKFZphtes3\_2h15.2)

(No Pfam data available for DKFZphtes3\_2h15.2)

DKF2phtes3\_2i5  
-----

group: testes derived

DKF2phtes3 2i5 encodes a novel 151 amino acid protein with weak similarity to. C.elegans  
cosmid F20D12.3

No informative BLAST results; No predictive prosite, pfam or SCOP motive.

The new protein can find application in studying the expression profile of testis-specific  
genes.

similarity to C.elegans F20D12.3

many ATGs in front of the start of the ORF,  
unspliced intron in 5' region?

Sequenced by EMBL

Locus: unknown

Insert length: 2142 bp

Poly A stretch at pos. 2121, polyadenylation signal at pos. 2102

```

1 GCAGTAAATA TGATATGAAA GAATTCTCTA ACTTGGGGGT GGCTTGTAAAC
51 CTGTAATAAA AATATTGCTA AAATACCTTC TCTCAGTTTG AAAAAGCATC
101 TGAGCAATCC TCAGTTATTG GTGAATTCCT ACCAGTGTCT AATCCTCTC
151 TTTCCGTAT GGTCTTAGTG TGGTTGTCTT GGTGTAGTAT TTCAAGAGGA
201 ACCGTGAGCA AGATGAAAAG AGAGTGGGAC TTGGAGCTAA GAACGTTTTT
251 GGCTTTAAGT GCTACGTTAA CTCATTAAAT TCTTAGTGAT CTTGGGGAAG
301 TCCCTTCACC AGTGTGAGCC TCAGTTTCTT TATCTAATAA GTAAGGATAA
351 TCTTACCCAC CTTATTGCGG GGGCCCGAGG ATTACATGAT TGGTGTAACA
401 GTAGCACCTT GTACATTGTA AAGGACTAAT ACCAGTGGAC TTAACTCTTG
451 GCTGGGCTTT GGAATTCCTG GTGGGACTTT TTAATCATGT AGATTCTCAG
501 GCCCTGCCTT GGCTGTGGA ACCACAGACT CTATAGGTGG GCCCTTCCAG
551 AAGGCTTCAT GGTGTGTTCT CATGTGGAAC CTGTGTTGCA AGCCACTGCA
601 TGGTGTACT GCTATTAAAC TTAATACTTA TATTTCTCTT ATTGTGTGGA
651 TATATCTGTG GTGTTTGCCC ATGTATACCT CATTTCATAT TTCTTAAAGA
701 ATAGAATGGA ATGGTTTAA GCACGCTACA TTGTCCAGGT TATACCCACA
751 GAAGAGCTGT TGTGTAACAG AATCAGCATC ATACCTGAAT CATTGTGTACA
801 TTGCATATAA GACTATGTCT AAGTAGAAGA TGCTATGAAA TCATGTCTGC
851 TGTGGGGCCA GGCATAATTA TGAATGTTAC TTAAGAGCAT AGGTGAGGTG
901 AGAAAAGGGA ATGTGACTAG TGTTTAGTA TTTCTTGGT GTGGGATGAA
951 GTATAATCTT TTTTTTTTTT TCTCAACAAA GCAGTAAAC TAGAAAGAG
1001 GAGAACTCTT CCTCAAGAA TGGCTGTACC TTCATATCTA GAGGCACATT
1051 AAAAAAAGA ACGTCTGTAC CTTAAAAATG GAGGTCATTT CATTGTGTTC
1101 ATTTTCAAGG TTGTTGTATG GCTCGGTGAG AACTTTCTGT TACCAGAAGA
1151 CACTCACATT CAGAATGCTC CATTTCAGT GTGTTTCACA TCTTTACGGA
1201 ATGGCGGCCA CCTGCATATA AAAATAAAAC TTAGTGGAGA GATCACTATA
1251 AATACTGATG ATATTGATTT GGCTGGTGAT ATCATCCAGT CAATGGCATC
1301 ATTTTGTGCT ATTGAAGACC TTCAAGTAGA AGCGGATTTT CCTGTCTATT
1351 TTGAGGAATT ACGAAAGGTG CTAGTTAAGG TGGATGAATA TCATTCACTG
1401 CATCAGAAGC TCAGTGCTGA TATGGCTGAT CATTCTAATT TGATCCGAAG
1451 TTTGCTGGTC GGAGCTGAGG ATGCTCGTCT GATGAGGGAC ATGAAAACAA
1501 TGAAGAGTCG TTATATGGAA CTCTATGACC TTAATAGAGA CTTGCTAAAT
1551 GGATATAAAA TTCGCTGTAA CAATCACACA GAGCTGTGG GAAACCTCAA
1601 AGCAGTAAAT CAAGCAATTC AAAGAGCAGG TCGTCTGCGG GTTGGAAAAC
1651 CAAAGAACCA GGTGATCACT GCTTGTGCGG ATGCAATTCG AAGCAATAAC
1701 ATCAACACAC TGTTCAAAAT CATGCGAGTG GGGACAGCTT CTTCTAGGT
1751 GAGGAAAATA CAGGTCATGA AGTTCCTGGC AAAGATTTTC TGTTAAAAAC
1801 CTATGCTGGT TTGCTTTGGA TCACACCTGT GTGAACCCCG GGTGCTAAGA
1851 ATGAAAATAA CCTTGGTGAG TTGTACAAAT TAAAGACAAA GAACCTACATG
1901 TGAAGATAGA CTTGCTTTCT ATTTTAAAT CAGTAGTAGT ACTGTTGCTG
1951 AATAATACTA GGTTTTATG GAATAGGATG AATGCTTTTG AAGTATTAGG
2001 GCTTCAGAGT CCAATTTTGC TTATTATGTT TATATAATAA CATATTTTTT
2051 TCTTGAATTT GCAATTGAGT TTGTACTTTT CAAATAGATT ATCTACTTTT
2101 TCATTAAAT GTAAAGATGT TAAAAAATAA AAAAAAATAA AA

```

#### BLAST Results

No BLAST result

#### Medline entries

Peptide information for frame 3

```

1 MASFFAIEDL QVEADFPVVF EELRKVLVKV DEYHSVHQKL SADMDHNSL
51 IRSLLVGAED ARLMRDMKTM KSRYMELYDL NRDLLNGYKI RCNNHTELLG
101 NLKAVNQAIQ RAGRLRVGKP KNQVITACRD AIRSNNINTL FKMIRVGTA
151 S

```

Alert BLASTP hits for DKF2phtes3 2i5, frame 3

**HSPs:**

```
Query:   140 LFKIMRVGTA 149
          + KI++ G +
Sbjct:   689 ITKILQYGAS 698
```

Pedant information for DKFZphtes3\_2i5, frame 3

## Report for DKFZphtes3\_2i5.3

```
SEQ      MASFAIEDLQVEADFVYFEELRKVLVKVDEYHSVHQKLSADMAHDSNLISLLVGAEAD
PRD      cccceeehhhhhccccccccccccccccccccccccccccccccccccccccccccccccccc

SEQ      ARLMRDCTMTKSRMYELLYNLRDLNLGYYIKRCHNCHHTLGNLKAVNQAIQRAGRLRVGKP
PRD      hhhhhcchhhhhheecchhhhhheeeecchhhhhhhhhhhhhhhhhhhhhhhhhccccccc
```

(No Pfam data available for DKFZphtes3\_2i5.3)

DKFZphtes3\_2119

group: testes derived

DKFZphtes3\_2119 encodes a novel 166 amino acid protein without similarity to known proteins.

No informative BLAST results; No predictive prosite, pfam or SCOP motive.

The new protein can find application in studying the expression profile of testis-specific genes.

unknown

complete cDNA, complete cds, no EST hits

Sequenced by EMBL

Locus: unknown

Insert length: 1079 bp

Poly A stretch at pos. 1053, polyadenylation signal at pos. 1038

```
1 CCACAGGACA CACTGTTCCC AGGGCACAGA CACCCCTGGGC TTTGGTTGGG
51 TCTTGGCCTC CAGGTAGGGC CCTGTTGGGC AGCGGGCAGC AACTCCTGAG
101 ACACACTACTGT GATTCTTGGT GGTGGCTGTG GTAAAAAACC TGCAGGGCTA
151 GAGTTTGGGG TGAGATTGAG CAGTAACTGT GGCCTCTCCT AGTGACAGTA
201 TGTCACCTCCC ACTCCAGCA CGCATGCCCA CAGGCCACGG CCTCCACATC
251 ACAAAACCCCC CACCAAGTTG CCCATCTATG GAGCAGCTCC CATACGGCAG
301 GGTGAGGCTC TTACCTCCAC CTCCAGGGCA CAGACAGGGG GAGCTCTGTC
351 TCACGTGTAAG GCAATGAGGA GAGTTGAGGG CCCAGACCAG GCTAGGGGCC
401 ATCCCTTTC CCGAGCAGGC CTCAGGGAAG GACCAGCCCC ATTCCCCTCT
451 GACCTAGCTC TTAGCCCAGG AGCCTGCATA GGAAGAAAG GACAGACAGG
501 GCCTCCTTAC TGGCTGACAC TCAGGAGGGG CTGGGGCAAG AGAGCAGAGG
551 GAGCGCAGGG CCAGGCAGGG GCTGCTGAGG ATCCATGGGA GCTCAGGGTG
601 CACAAAGGGG CTGCCCTTCC TGGGCTGAGG GCAGCATCCC TATGGGAGCT
651 GAGAAAGTCC AATCCTGAGA TGGGACAGTG CTGCCAGGG GTGTGTGGCT
701 GGGCCCTGAC AACAGTCTCC CCAAAAGTGA CCACATCACC AGGCTCAGTT
751 CCAGGAAGGC TGAGAAAGTC CCAGTACACT GAGGATGCAC CTCAGTTACA
801 TAAAATAAAT GAAACTGGAG TACTAACGTA CAGTTTAAAG GTTATAGTTA
851 CTATTTTAT ATGATATACT AGTAATTTT GAATAGGGTA AACTTTAGGT
901 GTTTTGACAC CAAAAGAAAA CTACATGAGT TCATGCATGT GTTAAATTGC
951 TTTACTGTAG TAATCATTTA CATGTATATG TATATATGAA TATAATTATG
1001 GGCTCATTA ATTTAAATAT TATAAATAGG TGACAAAGAA TAAAGTTAAC
1051 TGGAAAAAAA AAAAAAAAAA AAAAAAAAAA
```

## BLAST Results

No BLAST result

## Medline entries

No Medline entry

## Peptide information for frame 1

ORF from 364 bp to 861 bp; peptide length: 166

Category: putative protein

Classification: no clue

```
1 MRRVEGPDQA RGHPLSRAGL REGPAPFSPD LGLSPGACIG KKGQTGPPYW
51 LTLRRGWGKR AEGAQQQAGA AEDPWELRVH KGAALPGLQA ASLWELRKS
101 PEMGQCCPGV CGWALTTVSP KVTTSPPGVP GRLRSAQYTE DAPQLHKINE
151 TGVLTYSCLKV IVTIFI
```

## BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphtes3\_2119, frame 1

No Alert BLASTP hits found

Pedant information for DKFZphtes3\_2119, frame 1

Report for DKFZphtes3\_2119.1

[LENGTH] 166  
[MW] 17691.35  
[pI] 9.54  
[KW] All\_Beta  
[KW] LOW\_COMPLEXITY 7.23 %

SEQ MRRVEGPDQARGHPLSRAGLREGPAPFPDGLSPGACIGKKGQTGPPYWLTLRRGWGKR  
SEG .....  
PRD ccc

SEQ AEGAQGQAGAAEDPWELRVHKGAAALPGLQAASLWELRKSNPEMGQCCPGVCGWALTTVSP  
SEG xxx  
PRD ccc

SEQ KVTTSFGSVPGRLRSAQYTEDAPQLHKINETGVLTYSCLKVIVTIFI  
SEG .....  
PRD ccc

(No Prosite data available for DKFZphtes3\_2119.1)

(No Pfam data available for DKFZphtes3\_2119.1)

DKFZphtes3\_2m18  
-----

group: nucleic acid management

DKFZphtes3\_2m18 encodes a novel amino acid protein, with similarity to mouse Dhml.

The protein seems to play a role in nucleotide metabolism, RNA metabolism, but also in DNA repair and cell cycle. The yeast homologue is a DNA strand exchange protein required for sporulation and homologous recombination.

The novel protein can find application as multifunctional nuclease / exoribonuclease.

nearly identical to mouse Dhml

complete cDNA, complete cds, start at Bp 42, EST hits

Sequenced by EMBL

Locus: unknown

Insert length: 3022 bp

Poly A stretch at pos. 3004, polyadenylation signal at pos. 2981

```
1 CTCGTCAGCC GGTGCGCCGC CGCCTCCAGC CGTGTGCCGC TATGGGAGTC
51 CCGGCGTCTT TCCGCTGGCT CAGCCGCAAG TACCCGTCCA TCATAGTCAA
101 CTGCGTGGAA GAGAAGCCAA AAGAATGCAA TGGTGTAAGG ATTCCAGTTG
151 ATGCCAGTAA ACCTAATCCA AATGATGTGG AGTTTGATAA TCTGTATTGT
201 GATATGAATG GAATCATCCA TCCCTGTACT CATCCTGAAG ACAAAACGAGC
251 ACCAAAAAAT GAAGATGAAA TGATGGTTGC AATTTTGTAG TACATTGACA
301 GACTTTTCAG TATTGTAAGA CCAAGAAGAC TTCTCTACAT GGCAATAGAT
351 GGAGTGGCAC CACGTGCTAA AATGAACCAG CAGCGTTCAA GGAGGTTTCT
401 GGCATCAAAA GAAGGAATGG AAGCAGCAGT CGAGAAGCAG CGAGTCAGGG
451 AAGAAATATT GGCAAAAGGT GGCTTCTTTC CTCCAGAAGA AATAAAAGAA
501 AGATTGTACA GCAACTGTAT TACACCAGGA ACTGAATTCA TGGACAATCT
551 TGCTAAATGC CTTGCTGATT ACATAGCTGA TCGTTTAAAT AATGACCCCTG
601 GGTGGAAAAA TTTGACAGTT ATTTTATCTG ATGCTAGTGC TCCTGGTGAA
651 GGAGAACATA AAATCATGGA TTACATTAGA AGCAGAAAGC CCCAGCCTAA
701 CCATGACCCA AATACTCATC ATTGTTTATG TGGAGCAGAT GCTGATCTCA
751 TTATGCTTGG CCTTGCCACA CATGAACCGA ACTTTACCAT TATTAGAGAA
801 GAATTCAAAC CAAACAAGCC CAAACCATGT GGTCTTTGTA ATCAGTTTGG
851 ACATGAGGTC AAAGATTGTG AAGGTTTGCC AAGAGAAAAG AAGGGAAAGC
901 ATGATGAAGT TGCCGATAGT CTTCTTGTG CAGAAGGAGA GTTTATCTTC
951 CTTGCGCTTA ATGTTCTTCG TGAGTATTTG GAAAGAGAAC TCACAATGGC
1001 CAGCCTACCA TTCACATTG ATGTTGAGAG GAGCATTGAT GACTGGGTTT
1051 TCATGTGCTT CTTGTGGGGA AATGACTTCC TCCCTCATTT GCCATCGTTA
1101 GAGATTAGGG AAAATGCAAT TGACCGTTTG GTTAACATAT AAAAAAATGT
1151 GGTACACAAA ACTGGGGGTT ACCTTACAGA AAGTGGTTAT GTCAATCTGC
1201 AAAGAGTACA GATGATCATG TTAGCAGTTG GTGAAGTTGA GGATAGCATT
1251 TTTAAAGAGA GAAAGGATGA TGAGGACAGT TTTAGAAGAC GACAGAAAGA
1301 AAAAAGAAAG AGAATGAAGA GAGATCAACC AGCTTTCAC TCTAGTGGAA
1351 TATTAATCC TCATGCCTTG GGTCAAGAA ATTCAACAGG TTCTCAAGTA
1401 GCCAGTAATC CGAGACAAGC AGCCTATGAA ATGAGGATGC AGAATAACTC
1451 TAGTCCTTCG ATATCTCCTA ATACGAGTTT CACATCTGAT GGCTCCCCGT
1501 CTCCATTAGG AGGAATTAA GCGAAAGCAG AAGACAGTGA CAGTGAACCT
1551 GAGCCAGAGG ATAATGTCAG GTTATGGGAA GCTGGCTGGA AGCAGCGGTA
1601 CTACAAGAAC AAAATTGATG TGGATGCAGC TGATGAGAAA TTCCGTCGGA
1651 AAGTTGTGCA GTCGTACGTT GAAGGACTTT GCTGGGTTCT TAGATATTAT
1701 TACCAGGGCT GTGCTTCCTG GAAGTGGTAT TATCCATTTC ATTATGCACC
1751 ATTTGCTTCA GACTTTGAAG GCATTGCAGA CATGCCATCT GATTTTGAGA
1801 AGGGTACGAA ACCGTTTAAA CCACTAGAAC AACTTATGGG GGTATTTCCA
1851 GCTGCAAGTG GTAATTTTCT ACCTCCATCA TGGCGGAAGC TCATGAGTGA
1901 TCCTGATTCT AGTATAATTG ACTTCTATCC TGAAGATTTT GCTATTGATT
1951 TGAATGGGAA GAAATATGCA TGGCAAGGTG TTGCTCTCTT GCCATTCTGT
2001 GATGAGCGAA GGCTACGAGC TGCCCTAGAA GAGGTATACC CAGACCTCAC
2051 TCCAGAAGAG ACCAGAAGAA ACAGCCTTGG AGGTGATGTC TTATTGTGG
2101 GGAAACATCA CCCACTCCAT GACTTCATTT TAGAGCTGTA CCAGACAGGT
2151 TCCACAGAGC CAGTGGAGGT ACCCCCTGAA CTATGTCATG GGATTCAAGG
2201 AAAGTTTTCT TTGGATGAAG AAGCCATTCT TCCAGATCAA ATAGTATGTT
2251 CTCCTGTTCC TATGTTAAGG GATCTGACAC AGAACACTGT AGTCAGTATT
2301 AATTTTAAAG ACCCACAGTT TGCTGAAGAT TACATTTTAA AAGCTGTAAT
2351 GCTTCCAGGA GCAAGAAAGC CAGCAGCAGT ACTGAAACCT AGTGACTGGG
2401 AAAAATCCAG CAATGGACGG CAGTGGGAAG CTCAGCTTGG CTTTAAACCT
2451 GACCGGAGGC CTGTGCACCT GGATCAGGCA GCCTTCAGGA CTTTGGGCCA
2501 TGTGATGCCA AGAGGCTCAG GAACTGGCAT TTACAGCAAT GCTGCACCAC
2551 CACCTGTGAC TTACCAGGGA AACTTATACA GGCCTGTTT GAGAGGACAA
2601 GCCCAGATTC CAAAACCTAT GTCAAAATATG AGGCCCCAGG ATTCTGGCG
2651 AGGTCCTCCT CCCCTTTTCC AGCAGCAAGG GTTTGACAGA GCGTGTGGGG
```

```

2701 CTGAACCTCT GCTCCCATGG AACCGGATGC TGCAAACCCA GAATGCAGCC
2751 TTCCAGCCAA ACCAGTACCA GATGCTAGCT GGGCCTGGTG GGTATCCACC
2801 CAGACGAGAT GATCGTGGAG GGAGACAGGG ATATCCAGA GAAGGAAGGA
2851 AATACCCTTT GCCACCACCC TCAGGAAGAT ACAATTGGAA TTAAGCTTTT
2901 GTAAAGCTTT CCCAAATCCT TTCATCATTC TACAGTTTIA TGCTATTTGT
2951 GGAAAGATTT CCTTCTCAAG TAGTAGTTTT TAATAAACT ACAGTACTTT
3001 GTGTAACAAA AAAAAAAAAA AA

```

## BLAST Results

No BLAST result

## Medline entries

95192042:  
 Characterization of cDNA encoding mouse homolog of fission yeast dhpl+  
 gene: structural  
 and functional conservation.

97361754:  
 Cloning and characterization of mouse Dhml cDNA, a functional homolog  
 of budding yeast  
 SEPl.

## Peptide information for frame 3

ORF from 42 bp to 2891 bp; peptide length: 950  
 Category: strong similarity to known protein

```

1  MGVPAFFRWL SRKYPPIIVN CVEEKPKECN GVKIPVDASK PNPNDVEFDN
51  LYLDNMNGIIH PCTHPEDKPA PKNEDEMMVA IFEYIDRLFS IVRPRRLLYM
101  AIDGVAPRAK MNQQRSSRRF ASKEGMEAAV EKQVRREEIL AKGGFLPPPEE
151  IKERFDSNCI TPGTEFMDNL AKCLRYIIAD RLNNDPGWKN LTVILSDASA
201  PGEGERHKIMD YIRQRQAQPN HDPNTHHCLC GADADLIMLG LATHEPNFTI
251  IREEFKPNKP KPCGLCNQFG HEVKDCEGLP REKKGKHDEL ADSLPCEAGE
301  FIFLRLNVLR EYLERELTMA SLPTTFDVER SIDDWVFMCF FVGNDFLPHL
351  PSLEIRENAI DRLVNIYKNV VHKTGGYLTE SGYVNLQVRVQ MIMLAVGEVE
401  DSIFKRRKDD EDSFRRQKE KRRMRKRDQP AFTPSGILTP HALGSRNSPG
451  SQVASNPRQA AYEMRMQNNN SPSISPNTSF TSDGSPSLG GIKRKAEDSD
501  SEPEPEDNVR LWEAGWKQRY YKNKFQVDAA DEKFRKVVQ SYVEGLCWVL
551  RYYYQGCASW KWWYYPFHYAP FASDFEGIAD MPSDFEKGTK PFKPLEQLMG
601  VFPAASGNFL PPSWRKLMSD PDSSIIDFYP EDFAIDLNGK KYAWQGVALL
651  PFVDERRLRA ALEEVYPDLT PEETRRNSLG GDVLFVGGKH PLHDFILELY
701  QTGSTPEVEV PPELCHGIQG KFSLEDEAIL PDQIVCSVPV MLRDLTQNTV
751  VSINFKDPQF AEDYIFKAVM LPGARKPAAV LKPSDWEKSS NGRQWKPLQG
801  FNRDRRPVHL DQAAFRTLGH VMFRGSGTGI YSNAAPPPVT YQGNLYRPLL
851  RGOAQIPKLM SNMRPQDSWR GPPPLFQQQR FDRGVGAEP LPPNRMLQTO
901  NAAFQPNQYQ MLAGPGGYPP RRDRGGRQG YPREGRKYPL PPPSGRYNWN

```

## BLASTP hits

No BLASTP hits available

## Alert BLASTP hits for DKFZphtes3\_2ml8, frame 3

PIR:I49635 mouse Dhml protein - mouse, N = 1, Score = 4765, P = 0

PIR:S43891 dhpl protein - fission yeast (Schizosaccharomyces pombe), N  
 = 3, Score = 1172, P = 2e-197

PIR:S20126 exoribonuclease RAT1 (EC 3.1.11.-) - yeast (Saccharomyces  
 cerevisiae), N = 2, Score = 1146, P = 3.8e-175

PIR:S72531 exonuclease II - fission yeast (Schizosaccharomyces pombe),  
 N = 4, Score = 622, P = 4.2e-125

>PIR:I49635 mouse Dhml protein - mouse  
 Length = 947

HSPs:



Score = 4765 (714.9 bits), Expect = 0.0e+00, P = 0.0e+00  
Identities = 884/930 (95%), Positives = 895/930 (96%)

```
Query:      1  MGVPAFFRWLSRKYPSSIIVNCVEEKPKCEKNGVKIPVDASKPNPNDFDNLVLDMNGIIH  60
            MGVPAFFRWLSRKYPSSIIVNCVEEKPKCEKNGVKIPVDASKPNPNDFDNLVLDMNGIIH  60
Sbjct:      1  MGVPAFFRWLSRKYPSSIIVNCVEEKPKCEKNGVKIPVDASKPNPNDFDNLVLDMNGIIH  60

Query:     61  PCTHPEDKPAPKNEDEMMVAIFEYIDRLFSIVRPRRLLYMAIDGVAPRAKMNQQRSSRRFR  120
            PCTHPEDKPAPKNEDEMMVAIFEYIDRLFIIVRPRRLLYMAIDGVAPRAKMNQQRSSRRFR  120
Sbjct:     61  PCTHPEDKPAPKNEDEMMVAIFEYIDRLFNIVRPRRLLYMAIDGVAPRAKMNQQRSSRRFR  120

Query:    121  ASKEGMEAAVEKQVRVEEILAKGGFLPPEEIKERFDSNCITPGTEFMDNLAKCLRYIAD  180
            A K GMEAAVEKQVRVEEILAKGGFLPPEEIKERFDSNCITPGTEFMDNLAKCLRYIAD  180
Sbjct:    121  AIKGGMEAAVEKQVRVEEILAKGGFLPPEEIKERFDSNCITPGTEFMDNLAKCLRYIAD  180

Query:    181  RLNNDPGWKNLTVILSDASAPGEGEHKIMDYIRRQRAQPN DPNTHHCLCGADADLIMLG  240
            RLNNDPGWKNLTVILSDASAPGEGEHKIMDYIRRQRAQPN DPNTHHCLCGADADLIMLG  240
Sbjct:    181  RLNNDPGWKNLTVILSDASAPGEGEHKIMDYIRRQRAQPNQDPNTHHCLCGADADLIMLG  240

Query:    241  LATHEPNFTIIREEFKPNKPKPCGLCNQFGHEVKDCEGLPREKKGKHDELADSLPCAEGE  300
            LATHEPNFTIIREEFKPNKPKPC LCNQFGHEVKDCEGLPREKKGKHDELADSLPCAEGE  300
Sbjct:    241  LATHEPNFTIIREEFKPNKPKPCALCNQFGHEVKDCEGLPREKKGKHDELADSLPCAEGE  300

Query:    301  FIFLRNLVREYLERELTMASLPFTFDVERSIDDWVFMCFVGNDFLPHLSLEIRENAI  360
            FIFLRNLVREYLERELTMASLPF FDVERS DDW FMCFFVGNDFLPHLSLEIRE AI  360
Sbjct:    301  FIFLRNLVREYLERELTMASLPFPFDVERSNDWFMCFFVGNDFLPHLSLEIREGAI  360

Query:    361  DRLVNIYKNVHKTGGYLTESGYVNLQVRQMIMLAVGEVEDSIFKKRKDDSDSFRRRQKE  420
            DRLVNIYKNVHKTGGYLTESGYVNLQVRQMIMLAVGEVEDSIFKKRKDDSDSFRRRQKE  420
Sbjct:    361  DRLVNIYKNVHKTGGYLTESGYVNLQVRQMIMLAVGEVEDSIFKKRKDDSDSFRRRQKE  420

Query:    421  KRRKMRDQPAFTPSGILTPhALGSRNPGSQVASNPRQAAYEMRMQNNSSPSISPTSF  480
            KRRKMRDQPAFTPSGILTPhALGSRNPG QVASNPRQAAYEMRMQ NSSPSISPTSF  480
Sbjct:    421  KRRKMRDQPAFTPSGILTPhALGSRNPGCQVASNPRQAAYEMRMQRNSSPSISPTSF  480

Query:    481  TSDGSPSLGGIKRKAEDSDSEPEPEDNVRLWEAGWKQRYKKNKFDVDAADEKFRRKVVQ  540
            SDGSPSLGGI+RKAEDSDSEPEPEDNVRLWEAGWKQRYKKNKFDVDAADEKFRRKVVQ  540
Sbjct:    481  ASDGSPSLGGIRKKAEDSDSEPEPEDNVRLWEAGWKQRYKKNKFDVDAADEKFRRKVVQ  540

Query:    541  SYVEGLCWVLRYYYQGCASWKWYPFHYAPFASDFEGIADMPSDFEKGTKPFKPLEQLMG  600
            SYVEGLCWVLRYYYQGCASWKW YPFHYAPFASDFEGIADM S+FEKGTKPFKPLEQLMG  600
Sbjct:    541  SYVEGLCWVLRYYYQGCASWKWLYPFHYAPFASDFEGIADMSEFEKGTKPFKPLEQLMG  600

Query:    601  VFPAASGNFLPSPWRKLMSPDSSIIDFYPEDFAIDLNGKKYAWQGVALLPFVDERRLRA  660
            VFPAASGNFLPP+WRKLMSPDSSIIDFYPEDFAIDLNGKKYAWQGVALLPFVDERRLRA  660
Sbjct:    601  VFPAASGNFLPPTWRKLMSPDSSIIDFYPEDFAIDLNGKKYAWQGVALLPFVDERRLRA  660

Query:    661  ALEEVYPDLTPEETRRNSLGGDVLVFGKHHPLHDFILELYQTGSTPEVPELCHGIQG  720
            ALEEVYPDLTPEE RRNSLGGDVLVFGK HPL DFILELYQTGSTPEV+VPELCHGIQG  720
Sbjct:    661  ALEEVYPDLTPEENRRNSLGGDVLVFGKLHPLRDFILELYQTGSTPEVDVPELCHGIQG  720

Query:    721  KFSLDEEAILPDQIVCSPVPMRLDLTQNTVVSINFKDPQFAEDYIFKAVMLPGARKPAV  780
            FSLDEEAILPDQ VCSVPVPMRLDLTQNT VSNFNDPQFAEDY+FKA MLPARKPA V  780
Sbjct:    721  TFSLDEEAILPDQTVCSVPVPMRLDLTQNTAVSNFNDPQFAEDYVFKAAMLPGARKPATV  780

Query:    781  LKPSDWEKSSNGRQWKPOLGFNRDRRPVHLDQAAFRTLGHVMPRGSGTGIYSNAAPPVPT  840
            LKP DWEKSSNGRQWKPOLGFNRDRRPVHLDQAAFRTLGHV PRGSGT +Y+N A P  840
Sbjct:    781  LKPGDWEKSSNGRQWKPOLGFNRDRRPVHLDQAAFRTLGHVTPRGSGTSVYTNALLPAN  840

Query:    841  YQGNLYRPLLRGQAQIPKLMNSMRPDQSWRGPPPLFQQQRFDGVAEPLLPWNRMILQTQ  900
            YQGN YRPLLRGQAQIPKLMNSMRP+DSWRGPPPLFQQ RF+R VGAEPPLLPWNRM+Q Q  900
Sbjct:    841  YQGNMYRPLLRGQAQIPKLMNSMRPKDSWRGPPPLFQQHFEASVGAEPPLLPWNRMILQNQ  900

Query:    901  NAAFQPNQYQMLAGPGGYPPRRDD-RGGRQ  929
            NAAFQPNQYQML GPGGYPPRRDD RGGRQ
Sbjct:    901  NAAFQPNQYQMLGGPGGYPPRRDDHRGGRQ  930
```

Pedant information for DKFZphtes3\_2m18, frame 3

Report for DKFZphtes3\_2m18.3

[LENGTH]	950
[MW]	108582.68
[pI]	7.26
[HOMOL]	PIR:I49635 mouse Dhml protein - mouse 0.0
[FUNCAT]	08.01 nuclear transport [S. cerevisiae, YOR048c] 1e-123
[FUNCAT]	04.01.04 rna processing [S. cerevisiae, YOR048c] 1e-123

```

[FUNCAT] 30.10 nuclear organization [S. cerevisiae, YOR048c] 1e-123
[FUNCAT] 01.03.16 polynucleotide degradation [S. cerevisiae, YGL173c] 3e-79
[FUNCAT] 30.03 organization of cytoplasm [S. cerevisiae, YGL173c] 3e-79
[FUNCAT] 03.22 cell cycle control and mitosis [S. cerevisiae, YGL173c] 3e-79
[PIRKW] nucleus 1e-126
[PIRKW] hydrolase 1e-122
[PIRKW] exoribonuclease 1e-122
[PROSITE] MYRISTYL 7
[PROSITE] AMIDATION 2
[PROSITE] CAMP_PHOSPHO_SITE 1
[PROSITE] CK2_PHOSPHO_SITE 12
[PROSITE] TYR_PHOSPHO_SITE 1
[PROSITE] GLYCOSAMINOGLYCAN 1
[PROSITE] PKC_PHOSPHO_SITE 8
[PROSITE] ASN_GLYCOSYLATION 4
[KW] TRANSMEMBRANE 1
[KW] LOW_COMPLEXITY 6.21 %

```

```

SEQ MGVPAFFRWLSRKYPISIVNCVEEKPKCNGVKIPVDASKPNPNDFDNLVLDMNGIIH
SEG .....
PRD cccchhhhhhhhhccceeeeeeeccccccccccccccccccccccccccccccccccccccceee
MEM .....

```

```

SEQ PCTHPEDKPAKNEDEMMVAIFEYIDRLFSIVRPRRLLYMAIDGVAPRAKMNQQRSRFR
SEG .....
PRD cccccccccccccchhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh
MEM .....

```

```

SEQ ASKEGMEAAVEKQVRVEEILAKGGFLPPEEIKERFDSNCITPGTEFMDNLAKCLRYIAD
SEG .....
PRD hhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh
MEM .....

```

```

SEQ RLNDPGWKNLTVILSDASAPGEGEHKIMDYIRRQRAQPNHDPNTHHCLCGADADLIMLG
SEG .....
PRD hccccccccceeeeeeeccccccccchhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh
MEM .....

```

```

SEQ LATHEPFTTIIREEFKPNKPKPGCLCNQFGHEVKDCEGLPREKKGKHDELADSLPCAEGE
SEG .....
PRD cccccccccccccccccccccccccceeeccccccccccccccccchhhhhhhhhhhhhhhhhhh
MEM .....

```

```

SEQ FIFLRLNVLEIRELTMASLPFTFDVERSIDDWVFMCFVGNDFLPHLPSLEIRENAI
SEG .....
PRD cchhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh
MEM .....MMMMMMMMMMMMMMMMMMMM.....

```

```

SEQ DRLVNIYKNVVKHTGGYLTESGYVNLQVRQIMLAVGEVEDSIFKKRKDDSDSFRRRQKE
SEG .....
PRD hhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh
MEM .....

```

```

SEQ KRKRMRDQPAFTPSGILTPHALGSRNSPGSQVASNPRQAAYEMRMQNNSSPSISPTSF
SEG .....
PRD hhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh
MEM .....

```

```

SEQ TSDGSPSPLGGIKRKAEDSDSEPEPEDNVRLWEAGWKQRYYNKFVDVDADEKFRKRVVQ
SEG .....
PRD cccccccccchhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh
MEM .....

```

```

SEQ SYVEGLCWVLRYYYQGCASWKWYYPFHYAPFASDFEGIADMPDSEKGTGKPKPLEQLMG
SEG .....
PRD hhhhhhhheeeeeccccccccccccccccccccccccccccccccccccccccchhhhhhh
MEM .....

```

```

SEQ VFPAASGNFLPPSWRKLMSPDSSIIDFYPEDFAIDLNGKKYANQGVALLPFVDERRLRA
SEG .....
PRD hccccccccccccccccccccccccceeeccccccccceeeccccccccccccccccchhhhh
MEM .....

```

```

SEQ ALEEVPDLTPEETRRNSLGGDVLFGVGHHPHLDHFILELYQTGSTPEVVPPELCHGIQG
SEG .....
PRD hhhhhccccchhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh
MEM .....

```

```

SEQ KFSLDEEAILPDQIVCSPVPMRLDLTQNTVVSINFKDPQFAEDYIFKAVMLPGARKPAAV
SEG .....

```

```

PRD      cccccceeeccccceccccccccccccccccccccccccchhhheccccccccce
MEM      .....

SEQ      LKPSDWEKSSNGRQWKPOLGFNRDRRPVHLDQAAFTLGHVMPRGSGTGIYSNAAPPVVT
SEG      .....
PRD      eccccccccccccccccccccccccccccccccccccchhhhhhhhhcccccccccccccccc
MEM      .....

SEQ      YQGNLYRPLLRGQAQIPKLMSNMRPQDSWRGPPPLFQQQRFDRGVGAEP LLPWNRMLQTQ
SEG      .....
PRD      cccccchhhhhccccchhhhhccccccccccccccccccccchhhhhccccccccccccchhhhhh
MEM      .....

SEQ      NAAFQPNQYQMLAGPGGYPPRRDRGRQGYPREGRKYPLPPPSGRYNWN
SEG      ..... xxxxxxxxxxxxxxxxxxxxxxxx .....
PRD      hcccccccccecccccccccccccccccccccccccccccccccccccccccccc
MEM      .....

```

## Prosites for DKFZphtes3\_2m18.3

PS000001	190->194	ASN_GLYCOSYLATION	PDOC00001
PS000001	247->251	ASN_GLYCOSYLATION	PDOC00001
PS000001	468->472	ASN_GLYCOSYLATION	PDOC00001
PS000001	477->481	ASN_GLYCOSYLATION	PDOC00001
PS000002	826->830	GLYCOSAMINOGLYCAN	PDOC00002
PS000004	675->679	CAMP_PHOSPHO_SITE	PDOC00004
PS000005	11->14	PKC_PHOSPHO_SITE	PDOC00005
PS000005	116->119	PKC_PHOSPHO_SITE	PDOC00005
PS000005	413->416	PKC_PHOSPHO_SITE	PDOC00005
PS000005	559->562	PKC_PHOSPHO_SITE	PDOC00005
PS000005	613->616	PKC_PHOSPHO_SITE	PDOC00005
PS000005	674->677	PKC_PHOSPHO_SITE	PDOC00005
PS000005	868->871	PKC_PHOSPHO_SITE	PDOC00005
PS000005	944->947	PKC_PHOSPHO_SITE	PDOC00005
PS000006	63->67	CK2_PHOSPHO_SITE	PDOC00006
PS000006	331->335	CK2_PHOSPHO_SITE	PDOC00006
PS000006	499->503	CK2_PHOSPHO_SITE	PDOC00006
PS000006	501->505	CK2_PHOSPHO_SITE	PDOC00006
PS000006	541->545	CK2_PHOSPHO_SITE	PDOC00006
PS000006	573->577	CK2_PHOSPHO_SITE	PDOC00006
PS000006	583->587	CK2_PHOSPHO_SITE	PDOC00006
PS000006	619->623	CK2_PHOSPHO_SITE	PDOC00006
PS000006	624->628	CK2_PHOSPHO_SITE	PDOC00006
PS000006	670->674	CK2_PHOSPHO_SITE	PDOC00006
PS000006	723->727	CK2_PHOSPHO_SITE	PDOC00006
PS000006	784->788	CK2_PHOSPHO_SITE	PDOC00006
PS000007	659->667	TYR_PHOSPHO_SITE	PDOC00007
PS000008	125->131	MYRISTYL	PDOC00008
PS000008	375->381	MYRISTYL	PDOC00008
PS000008	450->456	MYRISTYL	PDOC00008
PS000008	600->606	MYRISTYL	PDOC00008
PS000008	825->831	MYRISTYL	PDOC00008
PS000008	829->835	MYRISTYL	PDOC00008
PS000008	926->932	MYRISTYL	PDOC00008
PS000009	638->642	AMIDATION	PDOC00009
PS000009	934->938	AMIDATION	PDOC00009

(No Pfam data available for DKFZphtes3\_2m18.3)

DKFZphtes3\_2m20

-----

group: testes derived

DKFZphtes3\_2m20 encodes a novel 183 amino acid protein without similarity to known proteins.

No informative BLAST results; No predictive prosite, pfam or SCOP motive.

The new protein can find application in studying the expression profile of testis-specific genes.

group: unknown

DKFZphtes3\_2m20 encodes a novel

amino acid protein without similarity to known proteins.

No informative BLAST results; No predictive prosite, pfam or SCOP motive.

The new protein can find application in studying the expression profile of testis-specific genes.

unknown

EST hits are only from testis or uterus librarys  
remaining intron in3' UTR see EST-BLAST

Sequenced by EMBL

Locus: unknown

Insert length: 1341 bp

Poly A stretch at pos. 1320, polyadenylation signal at pos. 1300

```
1  GCAATCCAGG AGCTGAATGG TAACCTCTCC ACAAGCGAAA ACTGTTCTGTG
51  AATACAAGCA AAAGGCCCCC CAAGAGGACC CCTGATATGA TCCAGCAGCC
101 TCGGGCCCCG CTGGTGTGGG AGAAGGCTTC TGGTGAAGGA TTTGGCAAAA
151 CCGCCGCTAT TATACAGCTC GCTCCTAAAG CTCCTGTTGA CCTGTGTGAG
201 ACAGAGAAAC TGAGGGCAGC CTTCTTTGCA GTCCCGTTGG AAATGAGAGG
251 GTCCTTCCTG GTGCTGCTCC TGAGGGAATG CTTCCGAGAC CTGAGCTGGC
301 TGGCACTCAT CCATAGCGTC CGTGGGGAGG CGGGGCTGCT GGTGACGAGT
351 ATCGTCCCGA AGACCCCGTT TTTCTGGGCC ATGCACATCA CTGAGGCTCT
401 GCACAGAAAC ATGCAGGCTC TGTTTAGCAC CCTGGCTCAG GCGGAGGAGC
451 AGCAGCCCTA CCTGGAGGCT CCACCGTTAT CGCGGGGACT CGCTGTCTGG
501 CAGAGTACCA CCTGGGGGAT TATGGACACG CCTGGAACAG GTGTGGGTG
551 CTGGACAGGG TGGACACCTG GGCTGTGGTC ATGTTCAATTG ATTTTGGACA
601 GTTGGCCACC ATCCCTGTGC AGTCTCTGCG CCAGCTAGAC AGCGACGACT
651 TCTGGACCAT CCCACCCCTG ACTCAGCCAT TCATGCTGGA GAAAGACATT
701 TTGAGTTTCG ATGAGGTTGT CCATCGAATC CTCAAAGGGA AAATCACTGG
751 TGCTTTGAAC TCGGCGGTAA CTGCTCCTGC ATCTAACTTG GCTGTTGTCC
801 CTCCTACTCT GCCCTTGGGG TGTCTGCAGC AGGCTGCTGC CTAGGCCTGG
851 ACACATTGCA CATCTAAAG TTTGAAGAGT CTAATAACG GGGCTTCCCT
901 CAGCATGTTT CCTCTCTGT TTGCCACGGA TCCAGAGCCA CTGCCCCTGT
951 CTCTCTGTAC CCCTTTCACT CTTGAGGCTT GGGAGGTGAA AAAGGCCAGA
1001 CTGTGCCAGG GATTGATTCA ATTTTGCTTT TACTCCCAGC TTCCCTCTCA
1051 AAAGAGAGTG AAGTCTCATT TGTCTGTGTT CTCAGTTTCC CCAACTTGGC
1101 ATGAACATTG GAACCAACA TAGGAACATA CCATTAGGTT GAAAGCCTGA
1151 GGCAGCTGGG ATGGTCTTTC TTGTGTCTCT TCTTTGCACC CCAGAGCATG
1201 ATATAAGTGG TCCTAACAGA TTCTGGATAA TGGAGAAGCC CTCTGCTGGT
1251 TTTCTTGGCA TTCCATGTAG AATAGGTAGA GAATATTAA CCAATGAGCA
1301 AATAAATGTT GGCATGTTT ATGAAAAAAA AAAAAAAAAA A
```

## BLAST Results

-----

No BLAST result

## Medline entries

-----

No Medline entry

## Peptide information for frame 2

ORF from 479 bp to 841 bp; peptide length: 121  
 Category: questionable ORF  
 Classification: no clue

1 MRGTRCLAEY HLG DYGHAWN RCWVLD RVD T WAVVMFIDFG QLATIPVQSL  
 51 RQLDSDDFWT IPPLTQPFML EKDILSSYEV VHRILKGKIT GALNSAVTAP  
 101 ASNLAVVPPL LPLGCLQQA A

## BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphtes3\_2m20, frame 2

No Alert BLASTP hits found

## Peptide information for frame 3

ORF from 87 bp to 635 bp; peptide length: 183  
 Category: putative protein  
 Classification: no clue

1 MIQQPRAPLV LEKASGEGFG KTAIIQLAP KAPVDLCETE KLRAAFFAVP  
 51 LEMRGSFLVL LLRECFRDL S WLALIH SVRG EAGLLVTSIV PKTPFFWAMH  
 101 ITEALHQNMQ ALFSTLAQAE EQQPYLEAPP LCAGLAVWQS TTWGIMDTPG  
 151 TGVGCWTGWT PGLWSCSLIL DSWPPSLCSL CAS

## BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphtes3\_2m20, frame 3

No Alert BLASTP hits found

## Pedant information for DKFZphtes3\_2m20, frame 2

## Report for DKFZphtes3\_2m20.2

[LENGTH] 121  
 [MW] 13436.69  
 [pI] 5.81  
 [KW] Alpha\_Beta

SEQ MRGTRCLAEYHLGDYGHAWNRCWVLD RVD T WAVVMFIDFGQLATIPVQSLRQLDSDDFWT  
 PRD ccchhhhcc

SEQ IPPLTQPFMLEKDILSSYEVVHRILKGKITGALNSAVTAPASNLAVVPPL LPLGCLQQA  
 PRD cccccchhhhhccchhhhhhhhhccccchhhhhcccccccccccccccccccccccc

SEQ A  
 PRD C

(No Prosite data available for DKFZphtes3\_2m20.2)

(No Pfam data available for DKFZphtes3\_2m20.2)

## Pedant information for DKFZphtes3\_2m20, frame 3

## Report for DKFZphtes3\_2m20.3

[LENGTH] 183  
 [MW] 19971.49  
 [pI] 5.31  
 [KW] Alpha\_Beta

```

SEQ  MIQQPRAPLVLEKASGEGFGKTAIIQLAPKAPVDLCETEKLRAAFFAVPLEMRGSFLVL
PRD  ccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc
SEQ  LLRECFRDLSWLALIHSVRGEAGLLVTSIVPKTPFFWAMHITEALHQNMQALFSTLAQAE
PRD  hhhhhhcchhhhhhhhhhhcccccccccccccccccccccccccccccccccccccccc
SEQ  EQQPYLEAPPLCAGLAVWQSTTWGIMDTPGTGVGCWTGWTPGLWSCSLILDSWPPSLCSL
PRD  hhhcccccccccccccccccccccccccccccccccccccccccccccccccccccccc
SEQ  CAS
PRD  ccc

```

(No Prosite data available for DKFZphtes3\_2m20.3)

(No Pfam data available for DKFZphtes3\_2m20.3)

DKFZphtes3\_2n9

group: testes derived

DKFZphtes3\_2n9 encodes a novel 184 amino acid protein with very weak similarity to Homo sapiens PAC clone DJ0771P04 from 7q11.21-q11.23.

No informative BLAST results; No predictive prosite, pfam or SCOP motife.

The new protein can find application in studying the expression profile of testis-specific genes.

unknown

on genomic level encoded by HS1186N24, no splice pattern but EST matches

Sequenced by EMBL

Locus: unknown

Insert length: 1000 bp

Poly A stretch at pos. 988, polyadenylation signal at pos. 970

```
1 CAACCTTTTAA AAGATGTGAA TTGGACAGCC AGACTTGCTT ATTTGTCTGA
51 TATCTTCAGT ATTTTAAAT GATCTTAATG CTTCTATGCA AGGGAAGAAT
101 GCAACTTATT TTTCATGGC AGATAAAGT GAAGGACAAA AACAGAAGTT
151 AGAAGCTTGG AAAACAGAA TTTCTACAGA TTGTTATGAC ATGTTTCATA
201 ATTTAACAAC AATTATCAAT GAAGTAGGTA ATGATCTTGA TATTGCACAT
251 CTGCGAAAAG TTATCAGTGA ACATCTTACA AATTGTTAG AATGTTTTGA
301 ATTTTATTTT CCATCAAAAG AAGATCCACG CATAGGAAAT TTGTGGATCC
351 AAAATCCATT TCTTTCATCA AAAGATAACT TAAATTTAAC TGTAACTCTA
401 CAGGATAAGT TGTGAAGCT GGCTACCGAC GAAGGATTGA AAATCAGTTT
451 TGAAAATACA GCATCACITC CTTCAATTTG GATAAAAGCT AAAAATGACT
501 ATCTTGAGCT TGCTGAGATT GCTTTAAAT TGCTGCTTCT TTCCCTCA
551 ACATACCTCT GTGAGACCGG ATTCTCTACT TTAAGTGTTA TTAAACAAA
601 ACATAGAAAC AGTTTAAATA TACATTATCC CCTGAGGTAG CATTGTCAATC
651 AATCCAACCT AGATTAGACA AATTAACAAG CAAGAAGCAA GCTCACTTAT
701 CACATTAAAA GCTTTAAATA TTGATATGTA AGGTATTGGT TCAAAGTATG
751 CATATAAGCA TTGAGTGTGA GGAATTTGCT ATTTCACTTT AAACCTTCIG
801 TCTAGTTACA GTTATGGAAG TATGAGAAGT TATGAGTGAA ACAGCAATTT
851 TCTATATAAA TTGCCTATAT GTATATTTTC AATTAAGAAAT GTGTACAGTT
901 TTTATAATTC TATTTTCTCT CATATTTGTC GTATTTATTA AAATATAATT
951 TTAATCTCTG TGATTCTAAT ATTAAACAT TTGATCTTAA AAAAAAAAAA
```

## BLAST Results

Entry HS1186N24 from database EMBLNEW:  
Human DNA sequence \*\*\* SEQUENCING IN PROGRESS \*\*\* from clone 1186N24  
Score = 4921, P = 5.8e-215, identities = 989/992

## Medline entries

No Medline entry

## Peptide information for frame 2

ORF from 86 bp to 637 bp; peptide length: 184  
Category: similarity to unknown protein  
Classification: no clue

```
1 MQGKNATYFS MADKVEGQKQ KLEAWKNRIS TDCYDMFHNL TTIINEVGND
51 LDIAHLRKVI SEHLTNLLEC FEFYFPSKED PRIGNLWIQN PFLSSKDNLN
101 LTVTLQDKLL KLATDEGLKI SFENTASLPS FWIKAKNDYP ELAEIALKLL
151 LLFPSTYLCE TGFSTLSVIK TKHRNSLNIH YPLR
```

## BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphtes3\_2n9, frame 2

TREMBLNEW:AC004883\_3 gene: "WUGSC:H\_DJ0771P04.2"; Homo sapiens PAC clone DJ0771P04 from 7q11.21-q11.23, complete sequence., N = 1, Score = 94, P = 0.042

>TREMBLNEW:AC004883\_3 gene: "WUGSC:H\_DJ0771P04.2"; Homo sapiens PAC clone DJ0771P04 from 7q11.21-q11.23, complete sequence.  
Length = 533

HSPs:

Score = 94 (14.1 bits), Expect = 4.3e-02, P = 4.2e-02  
Identities = 39/177 (22%), Positives = 75/177 (42%)

```
Query: 1 MQGKNATYFSMADKVEGQKQKLEAWKNRISTDCYDMFHNLTIIINEVGNDLD-IAHLRKV 59
      +QG + M D + KL W+ ++ + F L + L+ I + ++
Sbjct: 354 LQGHSSQIVTQMYDLIRAFKAKLCLWETHLTRNNLAHFPTLKLASRNESDGLNYPKIAEL 413

Query: 60 ISEHLTNLLECFEFYFPSKEDPRIGNLWIQNPFLLSSKDNLNLTVTLQDKLLKLATDEGLK 119
      +E L + F+ Y + + + +PF + D+++ LQ +++ L + LK
Sbjct: 414 KTEFQKRLSD-FKLY---ESELTL---FSSPFSTKIDSVH--EELQMEVIDLQCNTVLK 463

Query: 120 ISFENTASLPSEFWIKAKNDYPXXXXXXXFPSTYLCETGFSTLSVIKTKHRNSL 177
      ++ +P F+ YP F STY+CE FS + + KTK+ + L
Sbjct: 464 TRYDKVG-IPFQYKYLWGSYPKYKHHCAKILSMFGSTYICEQLFSIMKLSKTKYCSQL 520
```

Pedant information for DKFZphtes3\_2n9, frame 2

Report for DKFZphtes3\_2n9.2

```
[LENGTH] 184
[MW] 21203.53
[PI] 6.52
[KW] Alpha_Beta
[KW] LOW_COMPLEXITY 6.52 %
```

```
SEQ MQGKNATYFSMADKVEGQKQKLEAWKNRISTDCYDMFHNLTIIINEVGNDLDIAHLRKVI
SEG .....
PRD cccccchhhhhhhhhhhhhhhhhhhhhcchhhhhccceccccccccchhhhhhhhh

SEQ SEHLTNLLECFEFYFPSKEDPRIGNLWIQNPFLLSSKDNLNLTVTLQDKLLKLATDEGLKI
SEG .....
PRD hhhhhhhhhhhhhccccccccccccceccccccccccccceehhhhhhhhhhhcccee

SEQ SFENTASLPSEFWIKAKNDYPELAELKLLLPSTYLCETGFSTLSVIKTKHRNSLNIH
SEG .....
PRD eccccccccccccccccchhhhhhhhhhhhhccccccccccccceccccccccccccceec

SEQ YPLR
SEG ....
PRD cccc
```

(No Prosite data available for DKFZphtes3\_2n9.2)

(No Pfam data available for DKFZphtes3\_2n9.2)



DKFZphtes3\_30f4

group: testes derived

DKFZphtes3\_30f4 encodes a novel 192 amino acid protein without similarity to known proteins.

No informative BLAST results; No predictive prosite, pfam or SCOP motive.

The new protein can find application in studying the expression profile of testis-specific genes.

unknown

Sequenced by LMU

Locus: /map="717.2-8 cR from top of Chr8 linkage group"

Insert length: 1388 bp

Poly A stretch at pos. 1330, polyadenylation signal at pos. 1310

```
1  CACTGAGCCC TCCTCAGATG GTTAGTGGCT TCCAACAGCC ATCAGGAGTG
51 TTTCTTGAAT GCCCCAGGTG TGGAGGACTT GGTCTGTGAC CACCTAGAAC
101 CCCAGAGCTG AACAGGAAGC CGTCCCTGCA GCAACAAGAG GGCTGGAAGG
151 GGGAGCTGCA GGCCACCCCTC GGCTCTCCCA CTGCTGGGGC GGTGATGTTT
201 GGGTGACATG TTTGAAAAAT ACTCTTAAAG ATACCAACTG TTCCCTTATA
251 TGGCTAATGG TTTGTGCAGC CACCAGCGAT GGCGGCCCTT ATTAGAGACC
301 AGGTTTGTTA AAACACCAAA TATTGCTGTC CACACTAGAC ATTAACCGGC
351 TTCAGAAAAG ATGGACACCT TTCCCCACGC TGTTTCGCTT CTTAACTTTG
401 GTCCAGCTTT AGCCACCACA CAGCGTGTGA GGGACTGCTG CTGCGGAGTC
451 AGCCTCGTTT GTCCCTCCGC CTCCCACCAG CATGCGCCGC TTCTGAGAGA
501 CACCAGCTCC CTGCCTCCAA GCCTGGTGCC ACAGGCCTGT CGTGAGGGAC
551 CCCTGCTTCC GAGAGCTCCT GGGGGGGTTC TGCCCTTCAC CACCTGGGAG
601 AGGTGTCTAG TCAGTTCCGA GTTGAACAAG GCCCGTGAC ACAGCATGTT
651 GGGGGGCCAG CCCAAAGTTC TTGTACCTC CTCATGCAA GCCAGCCATC
701 ACCCTCCGGC CAGAGCTCAA GGTGGCCCTT TGGCCAGCCC CTCCTTGGGT
751 CCTCCAGGAG GACTGAGCAC CCCTCCTAGC GGCATCCCTT GCCCTCCACA
801 GTGCTGCCAG GGGCAGCTCG CTCTGTGCCG TGGACTGAGA CCATCCCCTG
851 GTGACAGAAT GACCCGTTTG TTGGAATATC CTCGTTGCCA GAGAACTCC
901 CCAGGCATCT CGGAACGAAA CTATTTAGTT CCATTGTGAA CTGGCCACGG
951 GACAGCTTTT TATCAACTTA TTAAGTTGGA GCACCTGAAT CGCGCTTGCT
1001 GAGTTAGCAG TGGTGGTAAG CGTGTGTTAA ACACATAATG TTACGTTTTA
1051 GGAGAGAGAG GTCGTAAAGA AGTGTGCTGT CGCTCATGAC TCTCTCTAT
1101 TAGTTGGGTA ACAGTGGCCT CATGTTTGTG TCTGTGTGTA CACAGAGCCC
1151 TTAGGTTCTG CTCTGTTTCT TTGCCAGGTG AATGTTTGTG GCATGCGCTG
1201 CTGTCCGCGC CCCTCTGTCC TGCCGAGGGT TCAGCTGTGC GGCGCCCTGA
1251 TTCTCTCCAT GCACACAGAA CCTCCTTGTG TCTGTTCTCT TGTCTCTCTG
1301 TGGCTGACTC AATAAACTTT TCCCTCTGAC ATGAAAAAAA AAAAAAAAAG
1351 AAAAAAAAAG AAAAAAAAAG AAAAAAAAAG AAAAAAAAAG
```

## BLAST Results

Entry HS548358 from database EMBL:

human STS EST67250.

Score = 2126, P = 1.5e-89, identities = 444/472

Entry HS670351 from database EMBL:

human STS WI-18501.

Score = 2089, P = 7.1e-88, identities = 445/476

## Medline entries

No Medline entry

## Peptide information for frame 1

ORF from 361 bp to 936 bp; peptide length: 192

Category: putative protein

Classification: no clue

```

1 MDTFSHAVSL LNFGPALATT QVRDCCCGV SLVCPASASHQ HAPLLRDTSS
51 LPPSLVPQAC REGPLLPRAP GGVLPTTWE RCQFSSELNK ARAHSMGLAQ
101 PKVLVTSSCK ASHHPPARAQ GGPLASPSLG PPGGLSTPPS GIPCPQCCQ
151 GHVALCRGLR PSPGDRMTRL LEMPRCQRNS PGISERNYLV PL

```

## BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphtes3\_30f4, frame 1

No Alert BLASTP hits found

Pedant information for DKFZphtes3\_30f4, frame 1

## Report for DKFZphtes3\_30f4.1

```

[LENGTH]      192
[MW]           20281.56
[pI]           9.21
[BLOCKS]      BL01013C Oxysterol-binding protein family proteins
[KW]           All Alpha
[KW]           LOW_COMPLEXITY    10.94 %

SEQ  MDTFSHAVSLLNFGPALATTQVRDCCCGVSLVCPASASHQHAPLLRDTSSSLPPLVPQAC
SEG  .....
PRD  cccchhhheeeccccchhhhhhhccccceeeccccccccccccccccccccccccccccc

SEQ  REGPLLPRAPGGVLPTTWERCFSSSELNKAHSMGLAQPKVLVTSSCKASHHPPARAQ
SEG  .....
PRD  cccccccccccccccccccchhhhhhhhhhhccccceeeccccccccccccccccccccc

SEQ  GGPLASPSLPGGLSTPPSGIPCPQCCQGHVALCRGLRPSPGDRMTRLLEMPRCQRNS
SEG  xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx
PRD  cccccccccccccccccccccccccccccchhhhhhhhhccccccchhhhhccccccccccc

SEQ  PGISERNYLVPL
SEG  .....
PRD  cccccccccccc

```

(No Prosite data available for DKFZphtes3\_30f4.1)

(No Pfam data available for DKFZphtes3\_30f4.1)

DKFZphtes3\_35b4

-----

group: cell cycle

DKFZphtes3\_35b4 encodes a novel 1780 amino acid protein which is C-terminal identical to human M-phase phosphoprotein-1 (MPPL).

The novel protein contains a N-terminal Pfam kinesin motor domain and a ATP/CTP-binding site motif A (P-loop). MPPL is expressed and phosphorylated in the metaphase. Therefore the novel protein seems to be involved in the mitotic spindle during cell division.

The new protein can find application in modulation of the mitotic spindle.

"M-phase phosphoprotein-1" extension

motor protein

Sequenced by DKFZ

Locus: /map="750\_H\_1; 758\_H\_7; 759\_C\_9; 847\_D\_4; 906\_D\_1; 931\_D\_3; 944\_C\_1; 750\_G\_12; 800\_A\_11; 512.1 CR from top of Chr10 linkage group"

Insert length: 6284 bp

No poly A stretch found, no polyadenylation signal found

```
1 ATCCGAGTGC TGCTCGCGGG TCTGGCTAGT CAGGCGAAGT TTGCAGAATG
51 GAATCTAATT TTAATCAAGA GGGAGTACCT CGACCATCTT ATGTTTTTAG
101 TGCTGACCCA ATTGCAAGGC CTTCAGAAAT AAATTTTCGAT GGCATTAAAGC
151 TTGATCTGTC TCATGAATTT TCCTTAGTTG CTCCAAATAC TGAGGCAAAAC
201 AGTTTTCGAAT CTAAAGATTA TCTCCAGGTT TGTCTTCGAA TAAGACCATT
251 TACACAGTCA GAAAAAGAAC TTGAGTCTGA GGGCTGTGTG CATATTCTGG
301 ATTCACAGAC TGTTGTGCTG AAAGAGCCTC AATGCATCCT TGGTCGGTTA
351 AGTAAAAAAA GCTCAGGGCA GATGGCACAG AAATTCAGTT TTTCCAAGGT
401 TTTTGGCCCA GCAACTACAC AGAAGGAATT CTTTCAGGTT TGCATTATGC
451 AACCAAGTAA AGACCTCTTG AAAGGACAGA GTCGTCTGAT TTTTACTTAC
501 GGGCTAACCA ATTCAGGAAA AACATATACA TTTCAAGGGA CAGAAGAAAA
551 TATTGGCATT CTGCCTCGAA CTTTGAATGT ATTATTGAT AGTCTTCAAG
601 AAAGACTGTA TACAAAGATG AACCTTAAAC CACATAGATC CAGAGAATAC
651 TTAAGGTTAT CATCAGAAAC AGAGAAAGAA GAAATTGCTA GCAAAAGTGC
701 ATTGCTTCGG CAAATTAAAG AGGTACTGTG GCATATGAT AGTGATGATA
751 CTCCTTATGG AAGTTTAACT AACTCTTTGA ATATCTCAGA GTTTGAAGAA
801 TCCATAAAAG ATTATGAACA AGCCAACCTG AATATGGCTA ATAGTATAAA
851 ATTTTCTGTG TGGGTTTCTT TCTTTGAAAT TTACAATGAA TATATTTATG
901 ACTTATTGTT TCCTGTATCA TCTAAATCCC AAAAGAGAAA GATGCTGCGC
951 CTTTCCCAAG ACGTAAAGGG CTATCTTTT ATAAAGATC TACAATGGAT
1001 TCAAGTATCT GATTCCAAAG AAGCCTATAG ACTTTTAAAA CTAGGAATAA
1051 AGCACACAGAG TGTTCCTTC ACAAAATTGA ATATGCTTC CAGTAGAAGT
1101 CACAGCATAT TCACTGTATA AATATTACAG ATTGAAGATT CTGAAATGTC
1151 TCGTGTAATT CGAGTCAGTG AATTATCTTT ATGTGATCTT GCTGGTTCAG
1201 AACGAACATAT GAAGACACAG AATGAAGGTG AAAGGTTAAG AGAGACTGGG
1251 AATATCAACA CTTCTTTATT GACTCTGGGA AAGTGTTATTA ACGTCTTGAA
1301 GAATAGTGAA AAGTCAAAGT TTCAACAGCA TGTGCCTTTC CGGGAAAGTA
1351 AACTGACTCA CTATTTTCAA AGTTTTTTTA ATGGTAAAGG GAAAAATTTGT
1401 ATGATTGTCA ATATCAGCCA ATGTTATTTA GCCTATGATG AAACACTCAA
1451 TGTATTGAAG TTCTCCGCCA TTGCACAAAA AGTTTGTGTC CCAGACACTT
1501 TAAATTCTTC TCAAGATAAA TTATTTGGAC CTGTCAAATC TTCTCAAGAT
1551 GTATCACTAG ACAGTAATTC AAACAGTAAA ATATTAAATG TAAAAAGAGC
1601 CACCATTTCA TGGGAAAATA GTCTAGAAGA TTTGATGGAA GACGAGGATT
1651 TGGTTGAGGA GCTAGAAAAC GCTGAAGAAA CTCAAATGTG GGAACATAAA
1701 CTTCTTGATG AAGATCTAGA TAAACATTA GAGGAAAATA AGGCTTTTAT
1751 TAGCCACGAG GAGAAAAGAA AACTGTTGGA CTTAATAGAA GACTTGAAAA
1801 AAAAACTGAT AAATGAAAAA AAGGAAAAAT TAACCTTGGA ATTTAAAAAT
1851 CGAGAAGAAG TTACACAGGA GTTTACTCAG TATTGGGCTC AACGGGAAGC
1901 TGACTTTAAG GAGACTCTGC TTCAAGAACG AGAGATATTA GAAGAAAAATG
1951 CTGAACGTCG TTTGGCTATC TTCAAGGATT TGGTTGGTAA ATGTGACACT
2001 CGAGAAGAAG CAGCGAAAGA CATTTGTGCC ACAAAGTTG AAACCTGAAGA
2051 AGCTACTGCT TGTTTGAAC TAAAGTTTAA TCAAATTTAA GCTGAATTAG
2101 CTAAACCATA AGGAGAATTA ATCAAAACCA AAGAAGGATT AAAAAAGAGA
2151 GAAATGAAT CAGATTCATT GATTCAAGAG CTTGAGACAT CTAATAAGAA
2201 AATAATTACA CAGAAATCAA GAATTTAAGA ATTGATAAAT ATAATTGATC
2251 AAAAAAGAAG TACTATCAAC GAATTTTACA ACCTAAAGTC TCATATGGAA
2301 AACACATTTA AATGCAATGA CAAGGCTGAT ACATCTTCTT TAATAATAAA
2351 CAATAAATTG ATTTGTAATG AAACAGTTGA AGTACCTAAG GACAGCAAAAT
2401 CTAAAACTCTG TTCAGAAAGA AAAAGAGTAA ATGAAAATGA ACTTCAGCAA
2451 GATGAACAC CAGCAAGAA AAGGCTATAT CATGTTAGTT CAGCTATCAC
2501 TGAAGACCAA AAGAAAAGTG AAGAAGTGCG ACCGAACATT GCAGAAATTG
2551 AAGACATCAG AGTTTTACAA GAAAATAATG AAGGACTGAG AGCATTTTTA
```

```

2601 CTCACATTTG AGAATGAAGT TAAAAATGAA AAGGAAGAAA AAGCAGAAAT
2651 AAATAAACAG ATTGTTCAAT TTCAGCAGGA ACTTCTCTTT TCTGAAAAAA
2701 AGAATTTAAC TTTAAGTAAA GAGGTCCAAC AAAATCAGTC AAATTATGAT
2751 ATTGCAATTG CTGAATTACA TGTGCAGAAA AGTAAAAATC AAGAACAGGA
2801 GGAAAAGATC ATGAAATTGT CAAATGAGAT AGAACTGCT ACAGAAGCA
2851 TTACAAATAA TGTTTCACAA ATAAAAATTA TGCACACGAA ATAGACGAA
2901 CTACGTACTC TTGATTCAAT TTCTCAGATT TCAACATAG ATTTGCTCAA
2951 TCTCAGGGAT CTGTCAAATG GTTCTGAGGA GGATAATTG CCAATACAC
3001 AGTTAGACCT TTTAGGTAAT GATTATTTGG TAAGTAAGCA AGTTAAAGAA
3051 TATCGAATTC AAGAACCCAA TAGGAAAAAT TCTTTCCACT CTAGATTGA
3101 AGCTATTTGG GAAGAATGTA AAGAGATTGT GAAGGCTCT TCCAAAAAAA
3151 GTCATCAGAT TGAGGAATCG GAACAACAAA TTGAAAAATT GCAGGCAGAA
3201 GTAAAAAGCT ATAAGGATGA AAACAATAGA CTAAAGGAGA AGGAGCATAA
3251 AAACCAAGAT GACCTACTAA AAGAAAAAGA AACTCTTATA CAGCAGCTGA
3301 AAGAAGAATT GCAAGAAAAA AATGTTACTC TTGATGTTCA AATACAGCAT
3351 GTAGTTGAAG GAAAGAGAGC GCTTTCAGAA CTTACACAAG GTGTTACTTG
3401 CTATAAGGCA AAAATAAAGG AACTTGAAAC AATTTTAGAG ACTCAGAAAG
3451 TTGAACGTAG TCATTACGCC AAGTTAGAAC AAGACATTTT GGAAAAGGAA
3501 TCTATCATCT TAAAGCTAGA AAGAAATTTG AAGGAATTTT AAGAATCTCT
3551 TCAGGATTCT GTCAAAAACA CCAAGATTTT AAATGTAAAG GAATCAAGC
3601 TGAAAGAGA AATCACACAG TTAACAAATA ATTTGCAAGA TATGAAACAT
3651 TTACTTCAAT TAAAGAAGA AGAAGAAGAA ACCAACAGGC AAGAAACAGA
3701 AAAATTGAAA GAGGAATCTC CTGCAACCTC TGCTCGTACC CAGAATCTGA
3751 AAGCAGATCT TCAGAGGAAG GAAGAAGATT ATGCTGACCT GAAGAGAGAA
3801 CTGACTGATG CCAAAAAGCA GATTAAAGCA GTACAGAAAG AGGTATCTGT
3851 AATCGCTGAT GAGGATAAAT TACTGAGGAT TAAATTAAT GAATCGGAGA
3901 AAAAGAAAAA CCAAGTGTCT CAGGAATTAG ATATGAAGCA GCCAACCAT
3951 CAGCAACTCA AGGAGCAGTT AAATAATCAG AAGTGGAG AAGCTATACA
4001 ACAGTATGAG AGAGCATGCA AAGATCTAAA TGTAAAGAG AAAATAATTG
4051 AAGACATGCG AATGACACTA GAAGAACAGG AACAACTCA GGTAGAACAG
4101 GATCAAGTGC TTGAGGCTAA ATTAGAGGAA GTTGAAGGC TGCCACAGA
4151 ATTTGAAAAA TGGAAAGAAA AATGCAATGA TTTGGAAACC AAAAAAATC
4201 AAAGGTCAAA TAAAGAACAT GAGAACAACA CAGATGTGCT TGGAAAGCTC
4251 ACTAATCTTC AAGATGAGTT ACAGGAGTCT GAACAGAAAT ATAATGCTGA
4301 TAGAAGAAAA TGGTTAGAAG AAAAAATGAT GCTTATCACT CAAGCGAAAG
4351 AAGCAGAGAA TATACGAAAT AAAGAGATGA AAAAAATGTC TGAGGACAGG
4401 GAGCGTTTTT TTAAGCAACA GAATGAAATG GAAATACTGA CAGCCAGCT
4451 GACAGAGAAA GATAGTGACC TTCAAAAGTG CCGAGAAGAA CGAGATCAAC
4501 TGGTTGACGC TTTAGAAATA CAGCTAAAAG CACTGATATC CAGTAATGTA
4551 CAGAAAGATA ATGAAATTGA ACAACTAAAA AGGATCATAT CAGAGACTTC
4601 TAAATAGAAA ACACAAATCA TGGATATCAA GCCCAACAGT ATTAGTTGAG
4651 CAGACTCTGA CAACTTCAA ACTGAACCTC TATCGACAAG TTTTGAAATT
4701 TCCAGAAATA AAATAGAGGA TGGATCTGTA GTCCTTGACT CTTGTGAAGT
4751 GTCAACAGAA AATGATCAAA GCATCTGATT TCCAAAACCT GAGTTAGAGA
4801 TTCAATTAC ACCTTTACAG CCAACAAAAA TGGCAGTGAA ACACCTGGT
4851 TGATCCACAC CAGTGACAGT TGAGATTCCC AAGGCTCGGA AGAGGAAGAG
4901 TGAATGAAATG GAGGAGGACT TGGTGAAATG TGAAAAAAG AAGAATGCTA
4951 CACCCAGAAC TAATTGAAA TTCTCTATT CAGATGATAG AAATCTTCT
5001 GTCAAAAAGG AACAAAAGGT TCCCATACGT CCATCATCTA AGAAAACATA
5051 TTCTTTACGG AGTCAGGCAT CCATAATTGG TGTAAACCTG GCCACTAAGA
5101 AAAAGAAGG AACACTACAG AAATTTGGAG ACTTCTTACA ACATTCTCCC
5151 TCAATTCTTC AATCAAAAGC AAAGAAGATA ATTGAAACAA TGAGCTCTTC
5201 AAAGCTCTCA AATGTAGAAG CAAGTAAAGA AAATGTGTCT CAACCAAAAC
5251 GAGCCAAACG GAAATTATAC ACAAGTGAAA TTTCACTCC TATTGATATA
5301 TCAGGCCAAG TGATTTTAAT GGACCAGAAA ATGAAGGAGA GTGATCAGCA
5351 GATTATCAAA CGACGACTTC GAACAAAAAC AGCCAAATAA ATCACTTATG
5401 GAAATGTTTA ATATAAATTT TATAGTCATA GTCATTGGAA CTTGCATCCT
5451 GTATTGTAAA TATAAATGTA TATATTATGC ATTAATACAC TCTGCATATA
5501 GATTGCTGTT TTATACATAG TATAATTTTA ATTCATAAAG TAGTCAAAA
5551 TTTGTATATT TTTATAAGGC TTTTATATA TAGCTTCTTT CAAACTGTAT
5601 TTCCCTATTA TCTCAGACAT TGGATCAGTG AAGATCCTAG GAAAGAGGCT
5651 GTTATTCTCA TTTATTTTGC TATACAGGAT GTAATAGGTC AGGTATTTGG
5701 TTTACTTATA TTTAACAATG TCTTATGAAT TTTTTTACT TTATCTGTTA
5751 TACAACATGAT TTTACATATC TGTTTGGATT ATAGCTAGGA TTTGGAGAAT
5801 AAGTGTGTAC AGATCAGAAA ACATGTATAT ACATTATTTA GAAAAGATCT
5851 CAAGTCTTTA ATTAGAATGT CTCATTATT TTGTAACAT TTTGTGGGTA
5901 CATAGTACAT GTATATATTT ACGGGGTATG TGAGATGTTT TGACACAGGC
5951 ATGCAATGTG AAATACGTGT ATCATGGAGA ATGAGGTATC CATCCCTCA
6001 AGCAATTTTC CTTTGAATTA CAGATAATCC AATTACATTC TTTAGATCAT
6051 TTAATAATAT ACAAGTAAGT TATTATTGAT TATAGTCACT CTATTGTGCT
6101 ATCAGATAGT AGATCATTCT TTTTATCTTA TTTGTTTTTG TACCCATTAA
6151 CCATCCCCAC CTCCCTGTC AACCGTCAGT ACCCTTACCA GCCACTGGTA
6201 ACCATTCTTC TACTCTGTAT GCCCATGAGG TCAATTGATT TTATTTTATG
6251 ATCCCATAAA TAAATGAGAA CATGCAAAAA AAAA

```

## BLAST Results

Entry HS898149 from database EMBL:  
human STS WI-9217.

Score = 4247, P = 1.5e-187, identities = 855/862

Medline entries

94119956:  
Cloning of cDNAs for M-phase phosphoproteins recognized  
by the MPM2 monoclonal antibody and determination of the  
phosphorylated epitope.

98101856:  
Interaction of a Golgi-associated kinesin-like protein with  
Rab6.

95122643:  
Identification and partial characterization of mitotic  
centromere-associated kinesin, a  
kinesin-related protein that associates with centromeres during  
mitosis.

Peptide information for frame 3

ORF from 48 bp to 5387 bp; peptide length: 1780  
Category: known protein  
Classification: Cell structure/motility  
Prosite motifs: ATP\_GTP\_A (152-160)

```
1 MESNFNQEGV PRPSYVFSAD PIARPSEINF DGIKLDLSHE FSLVAPNTEA
51 NSFESKDYLO VCLRIRPFTQ SEKELESEGC VHILDSQTVV LKEPQCILGR
101 LSEKSSGQMA QKFSFSKVFG PATTQKEFFQ GCIMQPVKDL LKGQSRILFT
151 YGLTNSGKTY TFQGTENIG ILPTLNLVLF DSLQERLYTK MNLKPHRSRE
201 YLRLSSEQEK EEIASKSALL RQIKVTVHN DSDDTLYGSL TNSLNISEFE
251 ESIKDYEQAN LNMANSIKFS VVVSFFFIYN EYIYDLFVPV SSKFQKRKML
301 RLSQDVKGYS FIKDLQWIVQ SDSKEAYRL KLGKHKQSV FTKLNNASSR
351 SHSIFTVKIL QIEDSEMSRV IRVSELSLCO LAGSERTMKT QNEGERLRET
401 GNINTSLTLL GKCINVLKNS ESKSFQOHVP FRESKLTHYF QSFNGKGI
451 CMIVNISQCY LAYDETLNVL KFSIAQKVC VPDTLNSQD KLFQPVKSSQ
501 DVSLDSNSNS KILNVKRATI SWENSLEDLM EDEDLVEELE NAEETQNVET
551 KLLDEDLTKT LEENKAFISH EKKRLLDLI EDLKKKLINE KKEKLTLEFK
601 IREEVTQEFQ QYWAQREADF KETLLQEREI LEENAERRLA IFKDLVGKCD
651 TREAAAKDIC ATKVETEEAT ACLELKFNOI KAEAKTKGE LIKTEELKK
701 RENESDSLII ELETSSNKKII TQNRKIKELI NIIDQKEDTI NEFQNLKSHM
751 ENTFCNDKA DTSSLIINNK LICNETVEVP KDSKSKICSE RKRNVNELQ
801 QDEPPAKKGS IHVSSAIED QKKSEEVVPN IAEIEDIRVL QENNEGLRAF
851 LLTINELKLN EKEEKAELNK QIVHFQOELS LSEKKNLTL SKEVQIQSNY
901 DIAIAELHVQ KSKNQEQEEK IMKLSNEIET ATRISITNVS QIKLMHTKID
951 ELRTLDSVSQ ISNIDLLNLR DLSNGSEEDN LPNTQLDLLG NDYLVSKQVK
1001 EYRIQEPNRE NSFHSSIEAI WEECKEIVKA SSKKSHQIEE LEQQIEKLQA
1051 EVKGYKDENN RLKEKEHKNO DDLLEKETL IQQLKEELOE KNTLDVQIO
1101 HVVEGKRALS ELTQGVTCYK AKIKELETIL ETQKVERSIS AKLEQDILEK
1151 ESIIILKERN LKEFQEHLDQ SVKNTKDLNV KELKKEEIT QLTNNLQDMK
1201 HLLQLKEEEE ETNRQETEKL KEELSASSAR TQNLKADLQR KEEDYADLKE
1251 KLTDAKKQIK QVQKEVSVMR DEDKLLRIKI NELEKKKNQC SQELDMKQRT
1301 IQQLKEQLNN QKVEEAIQY ERACKDLNVK EKIIEDMRMT LEEQEQTQVE
1351 QDQVLEAKLE EVERLATELE KWKEKCNLE TKNNQRSNKE HENNTDVLGK
1401 LTNLQDELQE SEQKYNADRK KWLEEKMLI TQAKEAENIR NKEMKKYAE
1451 RERFFKQONE MEILTAQLTE KSDSLQKWE ERDQLVALE IQLKALISSN
1501 VOKDNEIEQL KRIISETSKI ETQIMDIKPK RISSADPKL QTEPLSTSFE
1551 TSRNKIEDGS VVLDSECVST ENDQSTRFPK PELEIQFTPL QPNKMAVKHP
1601 GCTTPVTVEI PKARKRSNE MEEDLVKCN KKNATPRTNL KFPISDDRNS
1651 SVKKEQKVAI RPSSKKTYSL RSQASIIQVN LATKKKEGTL QKFGDFLOHS
1701 PSILQSKAKK IETMSSSKL SNVEASKENV SQPKRAKRL YLSEISSPID
1751 ISGQVILMDQ KMKESDHQII KRLRRTKTAK
```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphtes3\_35b4, frame 3

TREMBL:U93121\_1 product: "M-phase phosphoprotein-1"; Human M-phase  
phosphoprotein-1 mRNA, partial cds., N = 1, Score = 3743, P = 0

PIR:A36881 MPM2-reactive phosphoprotein 1 - human (fragment), N = 2,  
Score = 2808, P = 2.5e-294

TREMBL:AF070672.1 product: "rabkinesin6"; Homo sapiens rabkinesin6  
mRNA, complete Cds., N = 2, Score = 680, P = 2.6e-99

>TREMBL:U93121.1 product: "M-phase phosphoprotein-1"; Human M-phase  
phosphoprotein-1 mRNA, partial cds.  
Length = 753

## HSPs:

Score = 3743 (561.6 bits), Expect = 0.0e+00, P = 0.0e+00  
Identities = 752/753 (99%), Positives = 753/753 (100%)

Query: 1028 VKASSKKSHQIEELEQQIEKLQAEVKGKDYDENNRLEKEHEKNQDDLLKEKETLIQQLKEE 1087  
VKASSKKSHQIEELEQQIEKLQAEVKGKDYDENNRLEKEHEKNQDDLLKEKETLIQQLKEE  
Sbjct: 1 VKASSKKSHQIEELEQQIEKLQAEVKGKDYDENNRLEKEHEKNQDDLLKEKETLIQQLKEE 60

Query: 1088 LQEKNVTLDVQIQHVVEGKRALSELTQGVTCYKAKIKELETILETQKVERSHSAKLEQDI 1147  
LQEKNVTLDVQIQHVVEGKRALSELTQGVTCYKAKIKELETILETQKVERSHSAKLEQDI  
Sbjct: 61 LQEKNVTLDVQIQHVVEGKRALSELTQGVTCYKAKIKELETILETQKVERSHSAKLEQDI 120

Query: 1148 LEKESIILKLERNLKEFQEHLDQSVKNTKDLNVKELKLKEEITQLTNNLQDMKHLQLKE 1207  
LEKESIILKLERNLKEFQEHLDQSVKNTKDLNVKELKLKEEITQLTNNLQDMKHLQLKE  
Sbjct: 121 LEKESIILKLERNLKEFQEHLDQSVKNTKDLNVKELKLKEEITQLTNNLQDMKHLQLKE 180

Query: 1208 EEEETNRQETEKLEELSSASSARTQNLKADLQRKEEDYADLKEKLTDAKKQIKQVQKEVS 1267  
EEEETNRQETEKLEELSSASSARTQNLKADLQRKEEDYADLKEKLTDAKKQIKQVQKEVS  
Sbjct: 181 EEEETNRQETEKLEELSSASSARTQNLKADLQRKEEDYADLKEKLTDAKKQIKQVQKEVS 240

Query: 1268 VMRDEKLLRIKINELEKKNQCSQELDMKQRTIQQLKEQLNNQKVEEAIQOYERACKDL 1327  
VMRDEKLLRIKINELEKKNQCSQELDMKQRTIQQLKEQLNNQKVEEAIQOYERACKDL  
Sbjct: 241 VMRDEKLLRIKINELEKKNQCSQELDMKQRTIQQLKEQLNNQKVEEAIQOYERACKDL 300

Query: 1328 NVKEKIIEDMRMTLEEQEQTQVEQDQVLEAKLEEVEERLATELEKWKKECNDLETKNNQRS 1387  
NVKEKIIEDMRMTLEEQEQTQVEQDQVLEAKLEEVEERLATELEKWKKECNDLETKNNQRS  
Sbjct: 301 NVKEKIIEDMRMTLEEQEQTQVEQDQVLEAKLEEVEERLATELEKWKKECNDLETKNNQRS 360

Query: 1388 NKEHENNTDVLGKLTNLQDELQSESEQYNADRKWLEKMMMLITQAKEAENIRNKEMKKY 1447  
NKEHENNTDVLGKLTNLQDELQSESEQYNADRKWLEKMMMLITQAKEAENIRNKEMKKY  
Sbjct: 361 NKEHENNTDVLGKLTNLQDELQSESEQYNADRKWLEKMMMLITQAKEAENIRNKEMKKY 420

Query: 1448 AEDRERFFKQONEMEILTAQLTEKSDSLQKWEERDQVLALEIQLKALISSNVQKDNEI 1507  
AEDRERFFKQONEMEILTAQLTEKSDSLQKWEERDQVLALEIQLKALISSNVQKDNEI  
Sbjct: 421 AEDRERFFKQONEMEILTAQLTEKSDSLQKWEERDQVLALEIQLKALISSNVQKDNEI 480

Query: 1508 EQLKRIISETSKIETQIMDIKPKRISSADPKLQTEPLSTSFESISRNKIEDGSVVLDSC 1567  
EQLKRIISETSKIETQIMDIKPKRISSADPKLQTEPLSTSFESISRNKIEDGSVVLDSC  
Sbjct: 481 EQLKRIISETSKIETQIMDIKPKRISSADPKLQTEPLSTSFESISRNKIEDGSVVLDSC 540

Query: 1568 VSTENDQSTRFPKPELEIQFTPLQPNKMAVKHFGCTTPVTVEIPKARKKSNEMEEDLVK 1627  
VSTENDQSTRFPKPELEIQFTPLQPNKMAVKHFGCTTPVTVEIPKARKKSNEMEEDLVK  
Sbjct: 541 VSTENDQSTRFPKPELEIQFTPLQPNKMAVKHFGCTTPVTVEIPKARKKSNEMEEDLVK 600

Query: 1628 CENKKNATPRTNLKFPISSDRNSSVKKEQKVAIRPSSKKTYSLSQASIIGVNLATKKKE 1687  
CENKKNATPRTNLKFPISSDRNSSVKKEQKVAIRPSSKKTYSLSQASIIGVNLATKKKE  
Sbjct: 601 CENKKNATPRTNLKFPISSDRNSSVKKEQKVAIRPSSKKTYSLSQASIIGVNLATKKKE 660

Query: 1688 GTLQKFGDFLQHSPIILQSKAKKIIETMSSSKLSNVEASKENVSQPKRAKRKLYTSEISS 1747  
GTLQKFGDFLQHSPIILQSKAKKIIETMSSSKLSNVEASKENVSQPKRAKRKLYTSEISS  
Sbjct: 661 GTLQKFGDFLQHSPIILQSKAKKIIETMSSSKLSNVEASKENVSQPKRAKRKLYTSEISS 720

Query: 1748 PIDISGQVILMDQMKESDHQIIKRLRLTKTAK 1780  
PIDISGQVILMDQMKESDHQIIKRLRLTKTAK  
Sbjct: 721 PIDISGQVILMDQMKESDHQIIKRLRLTKTAK 753

Score = 197 (29.6 bits), Expect = 2.1e-11, P = 2.1e-11  
Identities = 114/542 (21%), Positives = 253/542 (46%)

Query: 692 IKTKELKKRENESDSLIQELETSSNKKIITQNRKELINIIDQKEDTINEFQNLKSHM- 750  
+K + + E + I++L+ K +N R+KE + ++D + E + L +  
Sbjct: 1 VKASSKKSHQIEELEQQIEKLQAEVKGKDYDENNRLEKEHE--KNQDDLLKEKETLIQQLK 58

Query: 751 ENTFCNKDADTS-SLIINNKLICNETVEVPKDSKSKICSERKRVNENELQDDEPPAK-- 807  
E + N D ++ K +E + K+KI E + + E + + AK  
Sbjct: 59 EELQEKNVTLDVQIQHVVEGKRALSELTQGVTCYKAKI-KELETILETQKVERSHSAKLE 117

Query: 808 KGSIHVSSAITEDQKSEEVPRNIAE-IEDIRVLQENNEGLRAFLLLTIENELKNEK---- 862

Sbjct: 118 + + S I + ++ +E + ++ + +++ + L L+ + + N L++ K  
 QDILEKESIILKLERNLKEFQEHLDQSVKNTKDLNVKELKKEEITQLTNNLQDMKHLLO 177  
 Query: 863 --EEKAELNKQIVH-FQOELSLSEKKNLTLSKEVQQIQSNYDIAIAELHVQKSKNQEQEE 919  
 EE+ E N+Q ++ELS S + L ++Q+ + +Y A+L K K + ++  
 Sbjct: 178 LKEEEETNRQETEKLEELSASSARTQNLKADLQRKEEDY----ADL---KEKLTDAKK 230  
 Query: 920 KIMKLSNEIETATRSITNNVSIKLMHTKIDEL-RTLDSVSIQSNIDLLNLRDLSNGSEE 978  
 +I ++ E+ S+ + + KL+ KI+EL + + SQ +D+ R + E+  
 Sbjct: 231 QIKQVQKEV-----SVMRD--EDKLLRIKINELEKKKNQCSQ--ELDMKQ-RTIQQLEQ 280  
 Query: 979 DNLPTQLDLLGNDYLVSKQVKEYRIQEPNRENSFHSSIEAIWEECKEIVKASSKSHQI 1038  
 N N +++ Y + K+ ++E E+ ++E + E + K +++  
 Sbjct: 281 LN--NQKVEEAIQY--ERACKDLNVKEKIID-MRMTLEEQEQTVQEQDQVLEAKLEE 335  
 Query: 1039 EEEQQIEKLQAEVKGKYNRLKEKEHKNQDDLLKEKETLIQQLKEELQEKVNT--- 1094  
 E L ++EK + + + NN+ KEH+N D+L + L +L+E Q+ N  
 Sbjct: 336 ERLATELEKWKECNDLETNNQRSNKEHENNTDVLGKLTNLQDELQEQSKYNADRKW 395  
 Query: 1095 LDVQIQHVVEGKRA-----LSELTQGVTCYKAKIKELETILETQKVERSHSAKLEQDI 1147  
 L+ ++ + + K A + + + + + + E+ IL Q E+ + ++  
 Sbjct: 396 LEEKMMLITQAKEAENIRNKEMKKAEDRERFFKQQNEME-ILTAQLTEKDSDLQKWE- 453  
 Query: 1148 LEKESIILKLERNLKEFQEHLDQSVKNTKOLNVKELK-LKEEITQLTNNLQDMKHLQLK 1206  
 E++ ++ LE LK + +V+ KD ++LK + E +++ + D+K +  
 Sbjct: 454 -ERDQLVALEIQLKAL---ISSNQ--KDNEIEQLKRIISETSKIETQIMDIK--PKR 504  
 Query: 1207 EEEETNRQETEKLEELSASSARTQN 1233  
 + ++ +TE L S + ++  
 Sbjct: 505 ISSADPKLQTEPLSTSFEISRNKIED 531  
 Score = 186 (27.9 bits), Expect = 3.2e-10, P = 3.2e-10  
 Identities = 131/674 (19%), Positives = 294/674 (43%)  
 Query: 673 LELKFNQKAEAKTKGELIKT-KEELKKRENSDSLIQELTSNKKIITQNRIKELIN 731  
 L+ K ++ + +L K K LI+ KEEL+++ D IQ + + + Q +  
 Sbjct: 35 LKEKEHKNQDDLLKEKETLIQQLKEELQEKVNTLDVQIQHVVEGKRALSELTQGVTCYKA 94  
 Query: 732 IIDQKEDTINEFQNL-KSHMENTFKCNDKADTSSLIINNKLICNETVEVPKDSKSKICSE 790  
 I + E TI E Q + +SH + D + S+I+ + E E +DS  
 Sbjct: 95 KIKELE-TILETQKVERSHSAKLEQ--DILEKESIILKLERNLKEFQEHLDQDS---VKN 147  
 Query: 791 RKRVENNELQ-QDEPPAKKGSIHVSSAITEDQKKSEEV-RPNIAEI-EDIRVLQENNEGL 847  
 K +N EL+ ++E ++ + + ++ EE R ++ E++ + L  
 Sbjct: 148 TKNLQVLEKLEKEEITQLTNNLQDMKHLQLKLEEEETNRQETEKLEELSASSARTQNL 207  
 Query: 848 RAFLLTIENELKNEKEEKAELNKQIVHFQOELSLSEKKNLTLSKEVQQI-----QSNYDI 902  
 +A L E + + KE+ + KQI Q+E+S+ ++ L ++ ++ Q + ++  
 Sbjct: 208 KADLQRKEEDYADLKEKLTDAKKQIKQVQKEVSVMRDEKLLRIKINELEKKKNQCSQEL 267  
 Query: 903 AIAELHVQKSKNQEQEEKIMKLSNEIETATRSITNNVSIKLMHTKIDEL-RTLDSVSIQ 961  
 + + +Q+ K Q +K+ + + E A + + I+ M ++E +T Q+  
 Sbjct: 268 DMKQRTIQQLEQLNNQKVEEAIQYERACKDLNVKEKIIDMRMTLEEQEQTVQEQDQV 327  
 Query: 962 SNIDLLNLRDLSNGSEEDNLPNTQLDLLGNDYLVSKQVKEYRI--QEPNRENSFHSSIEA 1019  
 L + L+ E+ L+ N + + + N ++ S +  
 Sbjct: 328 LEAKLEEVEERLATELEKWKECNDLETNNQRSNKEHENNTDVLGKLTNLQDELQEQSK 387  
 Query: 1020 IWEECKEIVKASSKSHQIEELEQQIEKLQAEVKGKYNRLKEKEHKNQ--DOLLKEK 1077  
 + K+ ++ Q+E E K E+K Y ++ R +++++ + L EK  
 Sbjct: 388 YNADRKWKLEEKMLITQAKEAENIRNK---EMKKAEDRERFFKQQNEMEILTAQLTEK 444  
 Query: 1078 ETLIQQLKEELQEKVNTLDVQIQHVVEGKRALSELTQGVTCYKAKIKELETILETQKVER 1137  
 ++ +Q+ +EE + L++Q++ ++ + + ++ ++ET + K +R  
 Sbjct: 445 DSDLQKWRERDQLVALEIQLKALISSNVQKDNEIEQLKRIISETSKIETQIMDIKPKR 504  
 Query: 1138 SHSAKLEQDILEKESIILKLERNLKEFQEHLDQDS---VKNTKDLNVKELKKEEITQLT 1193  
 SA ++ E S ++ RN E + DS +N + + +L+ + T L  
 Sbjct: 505 ISSADPKLQTEPLSTSFEISRNKIEDGSVVLDSCVSTENDQSTRFPKPELEIQFTPLQ 564  
 Query: 1194 NNLQDMKH---LLQLEEEETNRQETEKLEEL-SASSARTQNLKADLQRKEEDYADLK 1249  
 N +KH + + + ++++++E+L + + + +L+ D +  
 Sbjct: 565 PNKMAVKHPGCTTPVTVKIPKARKRKSNEEEDLVKCNKKNATPRTNLKFPISDDRNSS 624  
 Query: 1250 EKLTDAKKQIKQVQKEVSVMRDEKLLRIKINELEKKKNQCSQEL-DMKQRTIQQLEQL 1308  
 K + K I+ K+ +R + + I +N KKK Q+ D Q + L+ +  
 Sbjct: 625 VK-KEQKVAIRPSSKKTYSLSQASI--IGVNLATKKEGTLQKFGDFLQHSFILQSKA 681  
 Query: 1309 NNQKVEEAIQYERACKDLNVKEKIIDMR 1338  
 +K+ E + + + + + KE + + R  
 Sbjct: 682 --KKIETMSSSKLSNVEAS-KENVSQPKR 708

Score = 165 (24.8 bits), Expect = 5.8e-08, P = 5.8e-08  
Identities = 140/626 (22%), Positives = 271/626 (43%)

Query: 536 VEELNAEETQNVETKLLDEDLTKLEENKAFISHEEKRKLLDLIEDLKKKLINEKKEK- 594  
+EELE E E K +D + L+E + H+ + LL E L +L E +EK  
Sbjct: 11 IEELQEQIEKLAQEVKGY-KDENNRLEKE---HKNQDDLLKEKETLIQQLKEELQEK 65

Query: 595 LTLEFKIREVT-----QFTQYWAQREADFKE--TLQEREILEENAERRLAIFKDLVG 647  
+TL+ +I+ V E TQ +A KE T+L+ +++ E + +L +D++  
Sbjct: 66 VTLDVQIQHVVEGKRALSELTOGVTCYKAKIKELETILETQKV-ERSHSKLE--QDILE 122

Query: 648 KCDT---REEAAKDICATKVETEEATACLELKFNQIKAEAKTKGELIKTKEELKKRENE 704  
K E K+ ++ + T L +K +K E+ + L K L+ +E E  
Sbjct: 123 KESIILKLERNLKEFEHLQDSVKNTKDLNVKELKLEETQLTNNLQDMKHLQLKEEE 182

Query: 705 SDSLIQELETSSNKKIITONQRIKELINIIDQKEDTINEFQNLKSHMENTFKCNDKADTSS 764  
++ QE E +++ + R + L + +KE+ + + + + K K + S  
Sbjct: 183 EETNRQETKLEKELSSASSARTONLQADLQKEEDYADLKEKLTDAKKQIKQVQK-EVSV 241

Query: 765 LIINNKLICNETVEVPKDSKSKICSEKRVNENELQDEPPAKKSGSIHVSSAITEDQKKS 824  
+ +KL+ + E+ K K CS+ + + +QQ+ + V AI + ++  
Sbjct: 242 MRDEKLLRIKINELEK--KKNQCSQELDMKQRTIQQLEQLNNQK--VEEAIQYERAC 297

Query: 825 EEVRPNIAEIEDIRVLQENNEGLRAFLTIENELKNEKEEKAELNKQIVHFQQLSLSEK 884  
+++ IED+R+ E E + + + L+ + EE L ++ +++ + E  
Sbjct: 298 KDLNVKELIIEDMRMTLEEQQEQ---VEQDQVLEAKLEEVEERLATELEKWEKCNDET 354

Query: 885 KNLTLSEVQIQSNYDIAIAELHVQSKNQEQEEKIMKLSNE-IETATRSITN-----N 938  
KN S + + ++N D+ + +L + + QE E+K + +E IT N  
Sbjct: 355 KNNQSRNK--EHENNTDV-LGKLTNLQDELQSEQKYNADRKWLEEKMLITQAKEAEN 411

Query: 939 VSQIKLMHTKIDELRTLDVSQISNIDL-LNLRD--LSNGSEEDNLPNTQLDLLGNDYLV 995  
+ + + D R +++ + L +D L EE + L++ +  
Sbjct: 412 IRNKEMKKAEDRERFFKQKNEMEILTAQLTEKSDQKWRERDQLVAALEIQLKALIS 471

Query: 996 SKQVKYRIQEPNRENSFHSSIEA-IWE-ECKEIVKASSKKSHQIELEQIEKLAQEVK 1053  
S K+ I++ R S S IE I + + K I A K Q E L E + +++  
Sbjct: 472 SNVQKDNEIEQLKRIISETSKIETQIMDIKPKRISSADPKL-QTEPLSTSFEISRNKIE 530

Query: 1054 GYKDNNRLKEKEHKNQDDLLKEKE-----TLIQQLEKEELQEKNVTLDVQIQHVVEGKRA 1108  
+ + + +Q + E T +Q K ++ T V ++ KR  
Sbjct: 531 DGSVVLDSCEVSTENDQSTRFPKPELEIQFTPLQPNKMAVHKPGCTPTVTVKIPKARRK 590

Query: 1109 LSELTQG-VTCYKAKIKELETILETQ-KVERSHSAKLEQDILEKES 1152  
+E+ + V C K T L+ +R+ S K EQ + + S  
Sbjct: 591 SNEMEEDLVKCNKKNATPRTNLKFPISDDRNSSVKKEQKVAIRPS 636

Score = 143 (21.5 bits), Expect = 1.3e-05, P = 1.3e-05  
Identities = 164/684 (23%), Positives = 304/684 (44%)

Query: 295 QRRKMLR-LSQDVKGYSFIKDLQWIOVSDSKEAYRLKLGKIKHQSVAFKLNASS----- 349  
+K +++ L +++ + D+Q V + K A L G+ +L  
Sbjct: 49 EKETLIQQLKEELQEKNVTLDVQIQHVVEGKRALSELTOGVTCYKAKIKELETILETQKV 108

Query: 350 -RSHSI-FTVKILQIEDSEMSRVIRVSELSLCLAGSERTMKTONEGE-RLRETGNINTS 406  
RSHS IL+ E + + E L S + K N E +L+E T+  
Sbjct: 109 ERSHSKLEQDILEKESIIKLERNLKEFQE-HLQDSVKNTKDLNVKELKLEETQLTN 167

Query: 407 LLTLGKCINVLNSEKSKFQHVFPRESKLTHYFQSFFNGKKGICMIVNISQCYLAYDET 466  
L K + LK E+ +Q + +L+ N K + + Y E  
Sbjct: 168 NLQDMKHLQLKEEEEETNRQETKLEKELSSASSARTONLQADL---QRKEEDYADLKEK 224

Query: 467 LNVLFKSAIAQKVCVPTLNSSQDKLFGPVKSSQDVSLSNSNKKILNVKRATISWENSL 526  
L K I Q V ++ +DKL +K ++ + N S+ L++K+ TI  
Sbjct: 225 LTDAK-KQIKQ-VQKEVSMRDEKLLR-IKINE-LEKKKNQCSQELDMKQRTIQQLEKQ 280

Query: 527 EDLMEEDLVEELNAEETQNVETKLLDEDLTKLEENKAFISHEEKRKLLDL-IEDLKK 585  
+ + E+ +++ E A + NV+ K++ ED+ TLEE + + E+ ++L+ +E+++  
Sbjct: 281 LNNQVVEEAIQYERACKDLNVKELI-EDMRMTLEEQQEQ--TQVEQDQVLEAKLEEVEER 337

Query: 586 KLIN-EK-KEKLT-LEFKIREVTQFTQYWAQREADFKETLLQEREILEE---NAERR 638  
EK KEK LE K + +E + K T LQ+ E+ E NA+R+  
Sbjct: 338 LATELEKWEKCNDETCKNNQSRNKEHEN---NTDVLGKLTNLQD-ELQSEQKYNADRK 393

Query: 639 LAIFKDLVKGKCDTREEAAKDICATKVETEEATACLELKFNQIKAEAKTKGELIKTKEEL 698  
+ + ++ T+ + A++I K E ++ E F Q + E+ +L + +L  
Sbjct: 394 KWLEEKMM--LITQAKEAENI-RNK-EMKKAEDRERFFKQ-QNEMEILTAQLTEKSDL 448

Query: 699 KKRENSDSLIQELETSSNKKIITON-QR---IKELINIIDQKEDTINEFQNLKSHMENTF 754  
+K E D L+ LE K +I+ N Q+ I++L II + + ++K ++



Sbjct: 449 QKWEERDQLVAALEIQLKALISSNVQKDNEIEQLKRIISETSKIETQIMDIKPKRISSA 508

Query: 755 KCNDKADTSSLIINNKLICN--ETVEVPKDSKSKICSERK---RVNENELQ-QDEP--PA 806  
DK T L + ++ N E V DS ++ +E R + EL+ Q P P

Sbjct: 509 D-PDKLQTEPLSTSFEISRNKIEDGSVVLDS-CEVSTENDQSTRFPKPELEIQFTPLQPN 566

Query: 807 KKGSIH--VSSAITEDQKKSEEVPRNIAEIEDIRVLQENNEGLRA---FLLTIENELKNE 861  
K H ++ +T K+ + + N E + ++ + N R F + + + +

Sbjct: 567 KMAVHPGCTTPVTVKIPKARKRKSNEEEDLVKCNKKNATPRTNLKFPISDDRNSSVK 626

Query: 862 KEEKAEL---NKQIVHFQQLSLSEKKNLTLSKEVQIQSNYDIAIAELHVQKSKNQEQE 918  
KE+K + +K+ + + S+ NL K+ +Q D + +SK ++

Sbjct: 627 KEQKVATRPSSKKTYSLSRQASIIQV-NLATKKKEGTQKFGDFLQHSPSILQSKAKKII 685

Query: 919 EKIM--KLSNEIETATRSITNNVSQIKLMHTKI--DELRT-LDSVSQISNID 965  
E + KLSN +E + NVSQ K K+ E+ + +D Q+ +D

Sbjct: 686 ETMSSSKLSN-VEASKE---NVSQPKRAKRKLYTSEISSPIDISGQVILMD 732

Score = 133 (20.0 bits), Expect = 1.6e-04, P = 1.6e-04  
Identities = 94/426 (22%), Positives = 188/426 (44%)

Query: 527 EDLM-EDDELVEELENAEETQNVETKLLDEDLDTLEENKAFISHEEKRLDOL-IEDLK 584  
+DL+ E E L+++L+ + +NV LD + +E +A + I++L+

Sbjct: 44 DLLKEKETLIQQLKEELQEKNV---LDVQIQHVVEGKRALSELGTQVTCYKAKIKELE 100

Query: 585 KKLINKEKELTLEFKIRIEVTQ-EFTQYWAQREA-DFKETLLQEREILEENAERRLAIF 642  
L +K E+ + K+ +++ + E +R +F+E L + ++ + L +

Sbjct: 101 TILETQKVER-SHSAKLEQDILEKESIILKLERNLKEFQEHLDQSVKNTKDLNVKELKL- 158

Query: 643 KDLVGKCDTREEAAKDICATKVETEEATACLELKFNOIKAEALAKTKGELIKTKEELKKRE 702  
K+ + + + K + K E EE + +K EL+ + K +L+++E

Sbjct: 159 KEETITQLTNNLQDMKHLQLKEEEETN---RQTEKLKEELSASSARTQNLKADLQKE 215

Query: 703 NESDSLQIELETSNKKIITQNRKIKELINIIDQK-EDTINEFQNLKSHMENTFKCNDKA- 760  
+ L ++L T KK I Q Q+ ++ D+ INE + K+ +

Sbjct: 216 EDYADLKEKL-TDAKKQIKQVQKEVSMRDEKLLRLKINELEKKKNQCSQELDMKQRTI 274

Query: 761 DTSSLIINNKLICNETVE---VPKDS--KSKICSE-RKRVNENE---LQDEPPAKKGS 810  
+NN+ + E ++ KD K KI + R + E E ++QD+ K

Sbjct: 275 QQLKEQLNNQKV-EAIIQYERACKDLNVKEKIIEDMRMTLEEQQEQVQVLEAKLE 333

Query: 811 IHVSSAITEDQKKSEEVPR-NIAEIEDIRVLQENNEGLRAFLTTIENELKNEKEKAELN 869  
V TE +K E+ + ENN + L +++EL+ E E+K +

Sbjct: 334 -EVERLATELEKWEKCNDELTKNNQRSNKEHENNTDVLGKLTNLQDELQ-ESEQYNAD 391

Query: 870 KQIVHFQQLSLSEKKNLTLSKEVQIQSNYDIAIAELHVQKSKNQEQEEKIMKLSNEIE 929  
++ ++++ L +T +KE + I++ + K E E+ K NE+E

Sbjct: 392 RK-KWLEEKMML-----ITQAKEAENIRNK-----EMKKYAEDRRERFFKQQNEME 435

Query: 930 TATRSITNNVSQIKLMHTKIDEL 952  
T +T S ++ + D+L

Sbjct: 436 ILTAQLTEKDSQLQKWEERDQL 458

Pedant information for DKFZphtes3\_35b4, frame 3

#### Report for DKFZphtes3\_35b4.3

[LENGTH] 1780  
[MW] 206176.77  
[pI] 5.60  
[HOMOL] TREMBL:U93121\_1 product: "M-phase phosphoprotein-1"; Human M-phase phosphoprotein-1 mRNA, partial cds. 0.0  
[FUNCAT] 30.10 nuclear organization [S. cerevisiae, YEL061c] 2e-37  
[FUNCAT] 30.04 organization of cytoskeleton [S. cerevisiae, YEL061c] 2e-37  
[FUNCAT] 08.22 cytoskeleton-dependent transport [S. cerevisiae, YEL061c] 2e-37  
[FUNCAT] 03.22 cell cycle control and mitosis [S. cerevisiae, YEL061c] 2e-37  
[FUNCAT] 08.07 vesicular transport (golgi network, etc.) [S. cerevisiae, YDL058w] 7e-30  
[FUNCAT] 30.03 organization of cytoplasm [S. cerevisiae, YDL058w] 7e-30  
[FUNCAT] 30.05 organization of centrosome [S. cerevisiae, YPR141c] 3e-23  
[FUNCAT] 11.01 stress response [S. cerevisiae, YPR141c] 3e-23  
[FUNCAT] 03.07 pheromone response, mating-type determination, sex-specific proteins [S. cerevisiae, YPR141c] 3e-23  
[FUNCAT] 03.13 meiosis [S. cerevisiae, YPR141c] 3e-23  
[FUNCAT] 06.10 assembly of protein complexes [S. cerevisiae, YPR141c] 3e-23  
[FUNCAT] 09.10 nuclear biogenesis [S. cerevisiae, YPR141c] 3e-23  
[FUNCAT] 11.04 dna repair (direct repair, base excision repair and nucleotide excision repair) [S. cerevisiae, YKR095w] 1e-21

[FUNCAT] 99 unclassified proteins [S. cerevisiae, YLR309c] 6e-20  
 [FUNCAT] 03.04 budding, cell polarity and filament formation [S. cerevisiae, YHR023w  
 MYO1 - myosin-1 isoform] 4e-19  
 [FUNCAT] 03.25 cytokinesis [S. cerevisiae, YHR023w MYO1 - myosin-1 isoform] 4e-19  
 [FUNCAT] 03.19 recombination and dna repair [S. cerevisiae, YNL250w] 1e-15  
 [FUNCAT] 1 genome replication, transcription, recombination and repair [M.  
 jannaschii, MJ1322] 2e-14  
 [FUNCAT] 30.13 organization of chromosome structure [S. cerevisiae, YDR285w] 2e-09  
 [FUNCAT] 09.04 biogenesis of cytoskeleton [S. cerevisiae, YKL179c] 3e-09  
 [FUNCAT] 09.13 biogenesis of chromosome structure [S. cerevisiae, YLR086w] 2e-07  
 [FUNCAT] 03.01 cell growth [S. cerevisiae, YNL079c] 2e-07  
 [FUNCAT] 08.99 other intracellular-transport activities [S. cerevisiae, YNL079c]  
 2e-07  
 [FUNCAT] 03.22.01 cell cycle check point proteins [S. cerevisiae, YGL086w] 1e-06  
 [FUNCAT] 10.05.99 other pheromone response activities [S. cerevisiae, YHR158c]  
 3e-06  
 [FUNCAT] 04.05.01.04 transcriptional control [S. cerevisiae, YDR217c] 4e-06  
 [FUNCAT] 98 classification not yet clear-cut [S. cerevisiae, YJR134c] 2e-05  
 [FUNCAT] 05.04 translation (initiation, elongation and termination) [S. cerevisiae,  
 YAL035w] 2e-04  
 [FUNCAT] r general function prediction [M. jannaschii, MJ1254] 0.001  
 [BLOCKS] BL00387A  
 [BLOCKS] BL00411H  
 [BLOCKS] BL00411G  
 [BLOCKS] BL00411F  
 [BLOCKS] BL00411E Kinesin motor domain proteins  
 [BLOCKS] BL00411D Kinesin motor domain proteins  
 [BLOCKS] BL00411C Kinesin motor domain proteins  
 [BLOCKS] BL00411B Kinesin motor domain proteins  
 [BLOCKS] BL00411A Kinesin motor domain proteins  
 [SCOP] d2kin.1 3.29.1.5.3 Kinesin [Rat (Rattus norvegicus)] 2e-68  
 [SCOP] d2tmab 1.105.4.1.1 Tropomyosin [rabbit (Oryctolagus cuniculus)] 4e-05  
 [SCOP] d3kar 3.29.1.5.4 Kinesin [Baker's yeast (Saccharomyce 2e-09  
 [EC] 3.6.1.32 Myosin ATPase 5e-25  
 [PIRKEW] nucleus 4e-27  
 [PIRKEW] phosphotransferase 3e-16  
 [PIRKEW] duplication 6e-20  
 [PIRKEW] citrulline 6e-18  
 [PIRKEW] tandem repeat 4e-24  
 [PIRKEW] heterodimer 3e-28  
 [PIRKEW] endocytosis 1e-23  
 [PIRKEW] heart 1e-17  
 [PIRKEW] transmembrane protein 2e-28  
 [PIRKEW] serine/threonine-specific protein kinase 3e-16  
 [PIRKEW] zinc finger 1e-23  
 [PIRKEW] surface antigen 2e-16  
 [PIRKEW] DNA binding 1e-25  
 [PIRKEW] metal binding 1e-23  
 [PIRKEW] muscle contraction 4e-24  
 [PIRKEW] heterotetramer 4e-24  
 [PIRKEW] acetylated amino end 2e-19  
 [PIRKEW] actin binding 5e-25  
 [PIRKEW] mitosis 3e-58  
 [PIRKEW] microtubule binding 3e-58  
 [PIRKEW] ATP 3e-58  
 [PIRKEW] thick filament 4e-24  
 [PIRKEW] phosphoprotein 9e-29  
 [PIRKEW] leucine zipper 1e-12  
 [PIRKEW] skeletal muscle 8e-24  
 [PIRKEW] disulfide bond 1e-12  
 [PIRKEW] heterotrimer 1e-29  
 [PIRKEW] calcium binding 6e-18  
 [PIRKEW] alternative splicing 4e-21  
 [PIRKEW] P-loop 2e-63  
 [PIRKEW] coiled coil 3e-58  
 [PIRKEW] heptad repeat 1e-25  
 [PIRKEW] methylated amino acid 4e-24  
 [PIRKEW] peripheral membrane protein 1e-23  
 [PIRKEW] dimer 1e-12  
 [PIRKEW] cardiac muscle 1e-17  
 [PIRKEW] hydrolase 5e-25  
 [PIRKEW] microtubule 6e-15  
 [PIRKEW] muscle 7e-23  
 [PIRKEW] membrane protein 6e-20  
 [PIRKEW] GTP binding 8e-22  
 [PIRKEW] EF hand 6e-18  
 [PIRKEW] cell division 1e-25  
 [PIRKEW] cytoskeleton 4e-24  
 [PIRKEW] hair 6e-18  
 [PIRKEW] Golgi apparatus 8e-24  
 [PIRKEW] calmodulin binding 1e-23

[SUPFAM] unassigned Ser/Thr or Tyr-specific protein kinases 3e-16  
 [SUPFAM] myosin motor domain homology 5e-25  
 [SUPFAM] alpha-actinin actin-binding domain homology 1e-13  
 [SUPFAM] kinesin-related protein KIP1 9e-27  
 [SUPFAM] kinesin-related protein CIN8 4e-36  
 [SUPFAM] kinesin heavy chain 4e-24  
 [SUPFAM] plectin 1e-13  
 [SUPFAM] trichohyalin 6e-18  
 [SUPFAM] kinesin-related protein KIF3 1e-29  
 [SUPFAM] kinesin-related protein KIF2 3e-20  
 [SUPFAM] ribosomal protein S10 homology 1e-13  
 [SUPFAM] giantin 8e-24  
 [SUPFAM] protein kinase homology 3e-16  
 [SUPFAM] protein kinase C zinc-binding repeat homology 2e-13  
 [SUPFAM] kinesin-related protein unc-104 8e-26  
 [SUPFAM] human early endosome antigen 1 1e-23  
 [SUPFAM] unassigned kinesin-related proteins 1e-28  
 [SUPFAM] Mycoplasma genitalium hypothetical protein MG218 4e-17  
 [SUPFAM] myosin heavy chain 5e-25  
 [SUPFAM] conserved hypothetical P115 protein 4e-20  
 [SUPFAM] centromere protein E 5e-24  
 [SUPFAM] calmodulin repeat homology 6e-18  
 [SUPFAM] kinesin-related protein KLP61F 1e-25  
 [SUPFAM] hypothetical protein MJ0914 3e-12  
 [SUPFAM] kinesin-related protein MKLP-1 2e-63  
 [SUPFAM] pleckstrin repeat homology 8e-26  
 [SUPFAM] hypothetical protein MJ1322 4e-13  
 [SUPFAM] kinesin-related protein KIF1B 3e-28  
 [SUPFAM] kinesin motor domain homology 2e-63  
 [SUPFAM] kinesin-related protein KLPA 7e-25  
 [SUPFAM] kinesin-related protein nodA 1e-12  
 [SUPFAM] kinesin-related protein Eg5 5e-30  
 [PROSITE] ATP\_GTP\_A 1  
 [PFAM] Kinesin motor domain  
 [KW] Irregular  
 [KW] 3D  
 [KW] LOW COMPLEXITY 7.53 %  
 [KW] COILED\_COIL 19.78 %

SEQ MESNFNQEGVPRPSYVFSADPIARPSEINFDGKLDLSHEFSLVAPNTEANSFESKDYLO  
 SEG .....  
 COILS .....  
 3kar- .....  
 SEQ VCLRIRPFTQSEKELESEGCVHILDSQTVVLKEPQCILGRLEKSSGQMAQKFSFSKVFG  
 SEG .....  
 COILS .....  
 3kar- .....  
 SEQ PATTQKEFFQGCIMQPVKDLLKGQSRIFTYGLTNSGKTYTFQGTEENIGILPRTLNVLF  
 SEG .....  
 COILS .....  
 3kar- .....  
 SEQ DSLQERLYTKMNLKPHRSREYLRSLSEKEKEIASKSALLRQIKEVTVHNDSDDTLYGSL  
 SEG .....  
 COILS .....  
 3kar- .....  
 SEQ TNSLNISEFEESIKDYEQANLNMANSIKFSVWVSFFEIYNEYIYDLFVPVSSKFQKRKML  
 SEG .....  
 COILS .....  
 3kar- .....EEEEEEEEETTEEEETTCC-----CCEE  
 SEQ RLSQDVKGYSFIKDLQWIVQSDSKEAYRLKLGIKHQSVAFTKLNNASSRSHSIFTVKIL  
 SEG .....  
 COILS .....  
 3kar- EEEETTTT-EEETTCCEEECCGGGHHHHHHHHHHHCCTTTTCHHHHHHCEEEEEEEEE  
 SEQ QIEDSEMSRVIRVSELSLCLAGSERTMKTQNEGERLRETGNINTSLTLGKGINVLKNS  
 SEG .....  
 COILS .....  
 3kar- E--EETTTTCEEEEEEEECCECCCC---CCCHHHHHHHHHHHHHHHHHHHHHHHH  
 SEQ EKSFKQHVPPFRESKLTHYFQSFNGKGKICMIVNISQCYLAYDETNLVLKFSIAIAQKVC  
 SEG .....  
 COILS .....  
 3kar- TTTT--TCCTTTTTHHHHHHGGGCTTTTEEEEEEECCGGGHHHHHHHHHHHHHH.....  
 SEQ VPTDLNSSQDKLFGPVKSSQDVSLDSNSNKKILNVKRATISWENSLEDLMEDEDLVEEIE

SEG .....XXXXXXXXXXXXXXXXXXXX  
COILS .....  
3kar- .....  
  
SEG NAEETQNVETKLLDEDLTKLEENKAFISHEEKRKLLDLIEDLKKKLINEKKEKLTLEFK  
SEG .....XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX  
COILS .....CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC  
3kar- .....  
  
SEG IREEVTQEFQYWAQREADFKETLLQEREILEENAERRLAIFKDLVGKCDTREAAKDIC  
SEG .....  
COILS CCCCCC .....  
3kar- .....  
  
SEG ATKVETEEATACLELKFNOIKAEAKTKGELIKTKEELKKRENESDSLIELETSNKKII  
SEG .....  
COILS .....CCCCCCCCCCCCCCCC  
3kar- .....  
  
SEG TQNQRKELINIIDQKEDTINEFQNLKSHMENTFKCNDKADTSSLIINNKLICNETVEVP  
SEG .....  
COILS CCCCCCCCCCCCCC .....  
3kar- .....  
  
SEG KDSKSKICSEKRVNENELQQDEPPAKKGSIHVSSAITEDQKKSEEVPRNIAETIEDIRVL  
SEG .....  
COILS .....CCCC  
3kar- .....  
  
SEG QENNEGLRAFLTTIENELKNEKEEKAELNKQIVHFQQELSLSEKKNLTLSKEVQQIQSNY  
SEG .....XXXXXXXXXXXXXXXXXXXX  
COILS CC .....  
3kar- .....  
  
SEG DIAIAELHVQSKNQEQEEKIMKLSNEIETATRSITNNVSQIKLMTKIDELRTLDSVSO  
SEG .....  
COILS .....CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC  
3kar- .....  
  
SEG ISNIDLLNRLDLSNGSEEDNLPNTQLDLLGNDYLVSKQVKEYRIQEPNRENSFHSSIEAI  
SEG .....  
COILS .....  
3kar- .....  
  
SEG WEECKEIVKASSKKSHQIELEQQIEKLQAEVKGKYNENRLKEKEHKNQDDLLKEKETL  
SEG .....XXXXXXXXXXXXXXXXXXXX  
COILS .....CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC .....  
3kar- .....  
  
SEG IQQLKEELQEKNVTLDOVQIQHVVEGKRALSELTOGVTYKAKIKELETILETQKVERSHS  
SEG .....  
COILS .....CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC .....  
3kar- .....  
  
SEG AKLEQDILEKESIILKLERNLKEFQEHLDQSVKNKDLNVKELKKEBITQLTNLQDMK  
SEG .....  
COILS CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC ..... CCCCCCCCCCCCCCCCCCCCCCCCCC  
3kar- .....  
  
SEG HLLQLKEEEEETNRQETEKLEELSASSARTQNLKADLQKEEDYADLKEKLTDAKKQIK  
SEG .....XXXXXXXXXXXXXXXXXXXX  
COILS CCCCC ..... CCCCCCCCCCCCCCCCCCCCCCCCCCCCCC  
3kar- .....  
  
SEG QVQKEVSMRDEDKLLRIKINELEKKKNQCSQELDMKQRTIQQLKEQLNNQKVVEAIQOY  
SEG .....  
COILS CCCCCCCCCC .....  
3kar- .....  
  
SEG ERACKDLNVKEKIIEDMRMTLEEQEQTVQEQVLEAKLEEVERLATELEKWKECNDLE  
SEG .....XXXXXXXXXXXXXXXXXXXX  
COILS .....CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC  
3kar- .....  
  
SEG TKNNQRSNKEHENNTDVLGKLTNLQDELQSEQKYNADRRKKWLEEKMLITQAKEAENIR  
SEG .....  
COILS CC .....  
3kar- .....  
  
SEG NKEMKKYAEDRERFFKQONEMEILTAQLTEKDSDLQKWREERDQVLALEIQKALISSN  
SEG .....  
3kar- .....

```

COILS .....
3kar- .....

SEQ VQKDNEIEQLKRIISETSKIETQIMDIKPKRISSADPKLQTEPLSTSFEISRNKIEDGS
SEG .....
COILS .....
3kar- .....

SEQ VVLDSCEVSTENDQSTRFPKPELEIQFTPLQPNKMAVKHPGCTTPVTVEIPKARKRSNE
SEG .....
COILS .....
3kar- .....

SEQ MEEDLVKCNKKKNATPRTNLKFPISDDRNSSVKKEQKVAIRPSSKITYSLRSQASIIGVN
SEG .....
COILS .....
3kar- .....

SEQ LATKKKEGTLOKFGDFLOHSPSILQSKAKKIIETMSSSKLSNVEASKENVSQPKRAKRL
SEG .....
COILS .....
3kar- .....

SEQ YTSEISSPIDISGQVILMDQMKESDHQIKRRLRTKTAK
SEG .....
COILS .....
3kar- .....

```

## Prosite for DKFZphtes3\_35b4.3

PS00017 152->160 ATP\_GTP\_A PDOC00017

## Pfam for DKFZphtes3\_35b4.3

```

HMM_NAME Kinesin motor domain
HMM *RCRPINeREindgcscvVQWpPwtGyktvnhghegds.....phks
R+RP+ + E++ + +V + +++++ + + + +
Query 64 RIRPFTQSEKELESEGCVHILDSQTIVLKEPQCILGRLEKSSGQMAQK 112
HMM FcFDHVFVWncTQedVYdtvAHPIVDDcFhGYNCTIFAYGQTGSGKTYTM
F+F +VF++++TQ++ + + + V+D+++G IF+YG T SGKTYT
Query 113 FFSKVFVGPATTQKEFFQGCIMQPVKDLLKGQSRIFTYGLTNSGKTYTF 162
HMM MGpggehPDHmGIIPrcCHDIFdrIdkfgekDhdFW.....
G +++GI+PR+++ +ED++ + + +
Query 163 QG---TEENIGILPRTLNVLFDSLQERL-YTKMNLKPHRSREYLRSLSE 207
HMM .....
Query 208 QEKEEIASKSALLRQIKEVTVHNDSDDTLYGSLTNSLNISEFEESIKDYE 257
HMM .....hvkCSYMEIYNeeIYDLLCPnP...qhMkpLnIHEHPN
+V +S++EIVNE+IYDL +P++ Q++K L++ + +
Query 258 QANLNMANSIKFSVWVSFFEIYNEYIYDLFVPVSSKFQRRKMLRLSQDVK 307
HMM MGpYVqGCTEfHVCsYeDachWIWqGnknRHVAaTnMNdSSRShtIFTI
+++++ V +A +++ +G K+ VA T++N SSRShtIFT+
Query 308 GYSFIKDLQWIQVSDSKEAYRLLLKLGIKHQSVAFTKLNNASSRShtIFTV 357
HMM HVeQrHk.qcdehvcHSKMNLVDLAGSERvnrTGAEGQRlKEGcNINqSL
++ Q + + + +S ++L DLAGSER+ +T+ EG RL+E +NIN SL
Query 358 KILQIEDSEMSRVIRVSELSLCLAGSERTMKTQNEGERLRETGNINTSL 407
HMM ttLGnVInaLaDgqTKYmYgghgHIPYRDSKLTWlLQDSLGGNcKtCMIA
+TLG++IN+L + + + +H+P+R+SKLT+ +Q + G +K CMI+
Query 408 LTLGKCINVLKNSE---KSKFQQHVFPRESKLTHYFQSFFNGRGKICMIV 454
HMM CIWpadWNYEETLSTLRYAdRAKnIkNkPQINEDPca*
+I+ + Y+ETL++L++ + A+++ + ++N+++++
Query 455 NISQCYLAYDETlnVLKFSaIAQKVCVPTLNSSQDK 491

```

DKF2phtes3\_35b5

-----

group: metabolism

DKF2phtes3\_35b5 encodes a novel 466 amino acid protein, with similarity to bovine accessory subunit for vacuolar ATPase and rat C7-1 protein.

The vacuolar proton-ATPase (V-ATPase) translocates protons into intracellular organelles or across the plasma membrane of specialized cells. The catalytic domain consists of a hexamer of 3 A subunits and 3 B subunits, plus accessory subunits C, D, and E. The rat homolog C7-1 seems to be enriched in aged adult rats in the frontal cortex.

The novel protein can find application in modulating the v-ATPase activity in endocytic and secretory organelles.

strong similarity to bovine vacuolar ATPase (EC 3.6.1.-) chain A

complete cDNA, complete cds potential start at Bp 8, EST hits matches perfect to I54197 hypothetical protein, but possess 186 aa additional at N-terminus

Sequenced by DKFZ

Locus: unknown

Insert length: 2043 bp

Poly A stretch at pos. 2033, polyadenylation signal at pos. 2012

```
1  GCGGCGCATG GCGACGGCTC GAGTGGCGAT GGGGCGCGG TGCGCCCAGG
51 CGCTCTGGCG CATGCCGTGG CTGCCGGTGT TTTGTGCTT GCGGCGCGCG
101 GCGGCGCGCG CAGCGCGGGA GCAGCAGGTC CCGCTGGTGC TGTGGTCGAG
151 TGACCGGGAC TTGTGGGCTC CTGCGGCCGA CACTCATGAA GGCCACATCA
201 CCAGCGACTT GCAGCTCTCT ACCTACTTAG ATCCCGCCCT GGAGCTGGGT
251 CCCAGGAATG TGCTGCTGTT CCTGCAGGAC AAGCTGAGCA TTGAGGATTT
301 CACAGCATAT GCGCGTGTGT TTGGAACAAA GCAGGACAGC GCCTTTTCTA
351 ACCTAGAGAA TGCCCTGGAC CTGGCCCCCT CCTCACTGGT GCTTCTCGCC
401 GTGCACTGGT ATGCAGTCAG CACTCTGACC ACTTACCTGC AGGAGAAGCT
451 CGGGGCGCAG CCCTTGCACTG TGGACCTGGC CACCCTCGCG GAGCTGAAGC
501 TCAATGCCAG CCTCCCTGCT CTGCTGCTCA TTCGCTGCC CTACACAGCC
551 AGCTCTGGTC TGATGGCACC CAGGGAAGTC CTCACAGGCA ACGATGAGGT
601 CATCGGGCAG GTCCTGAGCA CACTCAAGTC CGAAGATGTC CCATACACAG
651 CGGCCCTCAC AGCGGTCCGC CCTTCCAGGG TGCCCCGTGA TGTAGCCGTG
701 GTGGCGGAGG GCGTAGGTGC CAGCTGCTCA CAAAAACAGC CAGTATCACC
751 TGTGATCCAT CCTCCTGTGA GTTACAATGA CACCGCTCCC CGGATCCTGT
801 TCTGGGCCCC AAACCTTCTCT GTGGCGTACA AGGACCAGTG GGAGGACCTG
851 ACTCCCTCA CTTTGGGGT GCAGGAAGTC AACCTGACTG GCTCCTCTGT
901 GAATGACTCC TTTGCCAGGC TCTCACTGAC CTATGAACGA CTCCTTGGTA
951 CCACAGTGAC ATTCAAGTTC ATTCTGGCCA ACCGCTCTA CCCAGTGTCT
1001 GCGCGGCACT GGTTTACCAT GGAGCGCCTC GAAGTCCACA GCAATGGCTC
1051 CGTCGGCTAC TTCAATGCTT CCGAGGTCAC AGGGCCGAGC ATCTACTCCT
1101 TCCACTGCGA GTATGTGAGC AGCCTGAGCA AGAAGGGTAG TCTCTCTGTG
1151 GCGCGCACGC AGCCCTCTCC CTGGCAGATG ATGCTTCAGG ACTTCCAGAT
1201 CCAGGCTTTC AACGTAATGG GGGAGCAGTT CTCCTACGCC AGCGACTGTG
1251 CCAGCTTCTT CTCCCCGGGC ATCTGGATGG GGCTGCTCAC CTCCTGTGTC
1301 ATGCTCTTCA TCTTCACCTA TGGCCTGCAC ATGATCTTCA GCCTCAAGAC
1351 CATGGATCGC TTTGATGACC ACAAGGGCCC CACTATTCTT TTGACCCAGA
1401 TTGTGTGACC CTGTGCCAGT GGGGGGGTTG AGGTTGGGAC GGTGTCCGTG
1451 TTGTGTGCTT CCCACCTGTC AGCGCACTGG ACTGAAGAGC TTCCTCTTTC
1501 CTAAGTGCAGC ATGAAGTCTC AGCTCCCTCT AGCCATCTTT GCTCCCTCTT
1551 CAGCCCGCTG AGGAGCTTTC TTGGGCTGCC CCCATCTCTC CCACCAAGGT
1601 GTACATATTC TGGGTAGATG CTAGACCAAC CAGCTTCCCA GGGTCTGTCG
1651 CTGTGAGGCG TAAGGGACAT GAATCTAGG GTCTCTTTC TCCTATTATA
1701 TTCTTGTGGC TACATCATCC CTGGCTGTGG ATAGTGTCTT TGTGTAGCAA
1751 ATGCTCCCTC CTTAAGGTTA TAGGGCTCCC TGAGTTTGGG AGTGTGGAAG
1801 TACTACTTAA CTGTCTGTCC TGCTTGGCTG CCCTTATCTG TTTCTGGTGA
1851 TGTGTGTGCT ACAATAAGAA GTACACGGGT TTATTCTCTG GGCCTGAGAA
1901 GGAAGGGACC TCCACGACAG GTGGGCTGGG TGCGATCGCC GGCTGTTTGG
1951 CATGTTCCCA CCGGAGGTGC CGGGCAGGAG CATGGGGTGC TTGGTTGTTT
2001 CTTCTCTAAT AAAATAAACG CGGGTCGCCA TGCAAAAAAA AAA
```

## BLAST Results

-----

No BLAST result

## Medline entries

95014142:  
A novel accessory subunit for vacuolar H(+)-ATPase from chromaffin granules.

97215246:  
Identification of a rat brain gene associated with aging by PCR differential display method.

## Peptide information for frame 2

ORF from 8 bp to 1405 bp; peptide length: 466  
Category: strong similarity to known protein

```
1 MATARVRMGP RCAQALWRMP WLPVFLSLAA AAAAAAEQQ VPLVLWSSDR
51 DLWAPAADTH EGHITSDQLQ STYLDPALEL GPRNVLLFLQ DKLSIEDFTA
101 YGGVFGNKQD SAFSNLENAL DIAPSSLVLP AVDWYAVSTL TTYLQEKLGA
151 SPLHVDLATL RELKLNASLP ALLLIRLPYT ASSGLMAPRE VLTGNDEVIG
201 QVLSTLKSED VPYTAALTAV RPSRVARDVA VVAGGLGRQL LQKQPVSPVI
251 HPPVSYNDTA PRILFWAQNF SVAYKDQWED LPTLFCVQOE LNLGTFWNSD
301 SFARLSLTIE RLFGTITVTFK FILANRLYPV SARHWTMER LEVHNGSVA
351 YFNASQVTGP SIYFHCETV SLSKKGSLV VARTQPSWQ MMLQDFQIQ
401 FNVMGQFSY ASDCASFFSP GIWMGLLSTL FHLFIFTYGL HMILSLKTM
451 RFDDHKGPTI SLTQIV
```

## BLASTP hits

No BLASTP hits available

## Alert BLASTP hits for DKFZphtes3\_35b5, frame 2

TREMBL:AF035387\_1 gene: "C7-1"; product: "C7-1 protein"; Rattus norvegicus C7-1 protein (C7-1) mRNA, complete cds., N = 1, Score = 2088, P = 3.8e-216

PIR:A55116 vacuolar ATPase (EC 3.6.1.-) chain Ac45 - bovine, N = 1, Score = 2011, P = 5.5e-208

PIR:I54197 hypothetical protein - human, N = 1, Score = 1464, P = 5.1e-150

>TREMBL:AF035387\_1 gene: "C7-1"; product: "C7-1 protein"; Rattus norvegicus C7-1 protein (C7-1) mRNA, complete cds.  
Length = 463

## HSPs:

Score = 2088 (313.3 bits), Expect = 3.8e-216, P = 3.8e-216  
Identities = 408/463 (88%), Positives = 426/463 (92%)

```
Query: 4 ARVRMGPRCAQALWRMPWLPVFLSLAAAAAAAEQQVPLVLWSSDRDLWAPAADTHEGH 63
      +R+R G R A LW + LSL A AAA AAEQQVPLVLWSSDRDLWAP ADTHEGH
Sbjct: 8 SRIRTGTRWAPVLW-----LLSLVAVAAVAAAEQQVPLVLWSSDRDLWAPVADTHEGH 61

Query: 64 ITSDQLSTYLDPALELGPRNVLLFLQDKLSIEDFTAYGGVFGNKQDSAFSNLENALDLA 123
      ITSD+QLSTYLDPALELGPRNVLLFLQDKLSIEDFTAYGGVFGNKQDSAFSNLENALDLA
Sbjct: 62 ITSDMQLSTYLDPALELGPRNVLLFLQDKLSIEDFTAYGGVFGNKQDSAFSNLENALDLA 121

Query: 124 PSSLVLPVADWYAVSTLTITTYLQEKLGASPLHVDLATLRELKLNASLPALLLIRLPYTASS 183
      PSSLVLPVADWYA+STLTITTYLQEKLGASPLHVDLATL+ELKLNASLPALLLIRLPYTASS
Sbjct: 122 PSSLVLPVADWYAISTLTITTYLQEKLGASPLHVDLATLRELKLNASLPALLLIRLPYTASS 181

Query: 184 GLMAPREVLTGNDEVIGQVLTSLKSEDPYTAALTAVRPSRVARDVAVVAGGLGRQLLQK 243
      GLMAPREVLTGNDEVIGQVLTSLKSEDPYTAALTAVRPSRVARDVAVVAGGLGRQLLQ
Sbjct: 182 GLMAPREVLTGNDEVIGQVLTSLKSEDPYTAALTAVRPSRVARDVAVVAGGLGRQLLQ 241

Query: 244 QPVSPVIHPPVSYNDTAPRILFWAQNFVAYKDQWEDLPTLFCVQELNLGTFWNSDFA 303
      Q SP IHPPVSYNDTAPRILFWAQNFVAYKD+H+DLT LTFGV+ LNLGTFWNSDFA
Sbjct: 242 QVASPAIHPPVSYNDTAPRILFWAQNFVAYKDEKDLTSLTFGVENLNLGTFWNSDFA 301

Query: 304 RLSLTIERLFGTITVTFKILANRLYPVSARHWTMERLEHVSNGSVA+YFNASQVTGPSIY 363
      LSLTYE LFG TVTFKILA+R YPVSAR+WTMERLE+HSNGSVA+FN SQVTGPSIY
```

Sbjct: 302 MSLTYEPLFGATVTFKFIASRFYPVSARYWFTMERLEIHSNGSVAHFNVSQVTGPSIY 361  
Query: 364 SFHCEYVSSLSKKGSLLVARTQPSWQMMLQDFQIQAFNVMEQFSYASDCASFFSPGIW 423  
SFHCEYVSSLSKKGSLLV PS WQM L +FQIQAFNV GEQFSYASDCA FFSPGIW  
Sbjct: 362 SFHCEYVSSLSKKGSLLVNTV-PSLWQMTLHNFQIQAFNVTEQFSYASDCAGFFSPGIW 420  
Query: 424 MGLLTSLFMLFIFTYGLHMILSLKTMDFDDHKGPTISLTQIV 466  
MGLLT+LFLMFIFTYGLHMILSLKTMDFDD KGPTI+LTQIV  
Sbjct: 421 MGLLTSLFMLFIFTYGLHMILSLKTMDFDDHKGPTITLTQIV 463

## Pedant information for DKF2phtes3\_35b5, frame 2

## Report for DKF2phtes3\_35b5.2

[LENGTH] 466  
[MW] 51621.44  
[pI] 5.73  
[HOMOL] TREMBL:AF035387\_1 gene: "C7-1"; product: "C7-1 protein"; Rattus norvegicus C7-1  
protein (C7-1) mRNA, complete cds. 0.0  
[PIRKW] hydrolase 0.0  
[PROSITE] MYRISTYL 7  
[PROSITE] CAMP\_PHOSPHO\_SITE 1  
[PROSITE] CK2\_PHOSPHO\_SITE 7  
[PROSITE] TYR\_PHOSPHO\_SITE 1  
[PROSITE] PKC\_PHOSPHO\_SITE 8  
[PROSITE] ASN\_GLYCOSYLATION 7  
[KW] SIGNAL PEPTIDE 38  
[KW] TRANSMEMBRANE 1  
[KW] LOW\_COMPLEXITY 11.59 %

SEQ MATARVRMGPRCAQALWRMPWLPVFLSLAAAAAAAEQVPLVLWSSDRDLWAPAADTH  
SEG .....XXXXXXXXX  
PRD cccccccccchhhhhccccchhhhhhhhhhhhhhhhhcccccccccccccccccc  
MEM .....  
SEQ EGHITSDLQLSTYLDPALELGPRLVFLQDKLSIEDFTAYGGVFGNKQDSAFSLENAL  
SEG .....XXXXXXXXXXXXXXXXXXXX  
PRD cccccchhhhhccccccccccccccccccccccccccccccccccccchhhhhhhcc  
MEM .....  
SEQ DLAPSSLVLPADVWYAVSTLTLYLQEKLGASPLHVDLATLRELKLNASLPALLLIRLPYT  
SEG .....XXXXXXXXXXXXXXXXXXXX  
PRD cccccccccccccccccccccccccccccchhhhhhhhhhhhhhhhhcccccccccccc  
MEM .....  
SEQ ASSGLMAPREVLTGNDEVIGQVSTLKSSEDVPTAALTAVRPSRVARDVAVAGGLGRQL  
SEG .....XXXXXXXXXXXXXXXXXXXX  
PRD cccccccccccccccccchhhhhccccccccchhhhhhhccccccccccccccccchh  
MEM .....  
SEQ LQKQPVSPVIHPPVSYNDTAPRILFWAQNFSVAYKQWEDLTPLTFGVQELNLTGSFWND  
SEG .....  
PRD hhhhhcc  
MEM .....  
SEQ SFARLSLTYERLFGTTVTFKFIANRLYPVSARHWFTMERLEVHSNGSVAYFNASQVTGP  
SEG .....  
PRD hhhhhhhhhhhccccccccccccccccchhhhhhhhhhhcccccccccccccccccc  
MEM .....  
SEQ SIYSFHCEYVSSLSKKGSLLVARTQPSWQMMLQDFQIQAFNVMEQFSYASDCASFFSP  
SEG .....XXXXXXXXX  
PRD cccccccccccccccccccccccccchhhhhhhhhhhcccccccccccccccccc  
MEM .....MMMMM  
SEQ GIWMGLTSLFMLFIFTYGLHMILSLKTMDFDDHKGPTISLTQIV  
SEG .....  
PRD ccchhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhcccccccccccccccccc  
MEM MMM

## Prosites for DKF2phtes3\_35b5.2

PS00001	166->170	ASN_GLYCOSYLATION	PDOC00001
PS00001	257->261	ASN_GLYCOSYLATION	PDOC00001
PS00001	269->273	ASN_GLYCOSYLATION	PDOC00001



WO 01/12659

PCT/IB00/01496

PS00001	292->296	ASN_GLYCOSYLATION	PDOC00001
PS00001	299->303	ASN_GLYCOSYLATION	PDOC00001
PS00001	346->350	ASN_GLYCOSYLATION	PDOC00001
PS00001	353->357	ASN_GLYCOSYLATION	PDOC00001
PS00004	375->379	CAMP_PHOSPHO_SITE	PDOC00004
PS00005	3->6	PKC_PHOSPHO_SITE	PDOC00005
PS00005	48->51	PKC_PHOSPHO_SITE	PDOC00005
PS00005	159->162	PKC_PHOSPHO_SITE	PDOC00005
PS00005	205->208	PKC_PHOSPHO_SITE	PDOC00005
PS00005	318->321	PKC_PHOSPHO_SITE	PDOC00005
PS00005	331->334	PKC_PHOSPHO_SITE	PDOC00005
PS00005	374->377	PKC_PHOSPHO_SITE	PDOC00005
PS00005	445->448	PKC_PHOSPHO_SITE	PDOC00005
PS00006	48->52	CK2_PHOSPHO_SITE	PDOC00006
PS00006	72->76	CK2_PHOSPHO_SITE	PDOC00006
PS00006	94->98	CK2_PHOSPHO_SITE	PDOC00006
PS00006	114->118	CK2_PHOSPHO_SITE	PDOC00006
PS00006	159->163	CK2_PHOSPHO_SITE	PDOC00006
PS00006	193->197	CK2_PHOSPHO_SITE	PDOC00006
PS00006	255->259	CK2_PHOSPHO_SITE	PDOC00006
PS00007	207->214	TYR_PHOSPHO_SITE	PDOC00007
PS00008	102->108	MYRISTYL	PDOC00008
PS00008	103->109	MYRISTYL	PDOC00008
PS00008	200->206	MYRISTYL	PDOC00008
PS00008	295->301	MYRISTYL	PDOC00008
PS00008	314->320	MYRISTYL	PDOC00008
PS00008	421->427	MYRISTYL	PDOC00008
PS00008	425->431	MYRISTYL	PDOC00008

(No Pfam data available for DKFZphtes3\_35b5.2)

DKF2phtes3\_35e21

group: differentiation/development

DKF2phtes3\_35e21.2 encodes a novel 104 amino acid putative interleukin precursor, related to interleukin-7.

Due to the close relationship to human interleukin-7, the novel interleukin is expected to act as a new growth factor for human B lineage cells. Additionally, the protein should induce the gene rearrangement of the T-cell receptor repertoire, leading to thymocyte commitment, and subsequently induce both cytotoxic T-cell- and lymphocyte-activated killer cells.

This new interleukin could find clinical application in a variety of conditions of hematolymphopoietic failure and different tumours, because of its recruitment of B cell lineage cells, cytotoxic T-cell- and lymphocyte-activated killer cells.

similarity to interleukin-7 precursor

complete cDNA, complete cds, EST hits

Sequenced by DKFZ

Locus: unknown

Insert length: 2095 bp

Poly A stretch at pos. 2085, polyadenylation signal at pos. 2067

```
1 GGATGAAAGT GATTAAATC ATTTTAGAA TTTTTTTT GTTTGTTTT
51 AGCAACATGC TGAACAATA ATTACTTTA AAAATAAGCC AGTTAAACA
101 AAGGACGCTA AGCCCAAGT GGGGGCAATA TTAGTCAGGA TCTTTGGGGT
151 CTAATTCAG ACCAACTTC AGAAGCACT CTTGTCTCT GTTCTCACCT
201 CTGCTGCCCT TCTCTCCCT CATCCCCTA GAGAGACAA GATAAAGGCC
251 CACCTGCATC CCTAAGTCT ACTGAGATCA GCCACCCAG GGGAGAGAAA
301 CTGGATCTAC TTACAGCCAC CCCCTGTTT CATCCATATA CTTACTTCCC
351 CCAATTTGCA TGTGATTATG GAAACAAGTC ATGCTCATGA AAGCAACTGT
401 AAAATAAAG GTTATGGAGT AGTTACAGAA CTTCTTCACA GCCAGCTTTG
451 TGGAGCTGGG GAGGACTTAG GGGCCATTGG AGTCTCTTAT GTGTACAGCT
501 TCAGGCTGCT CCCTTTCAGT TTGATTTTAA GCAATGCCTC ACTTCATAGC
551 TTAGGGGGTA AGGATTCCAT TCAGGTAGGT TGTCTAAAGG AACTAATGGG
601 ACCTCTCAGT GAATTAGCTG ACCAGATTTT AGGAAATCCT TTTAATTTCT
651 ATGATTTTCC TTCTCACATT TTGAAATGGT AAAATTGACT GGAATAATTT
701 TTTCTTGGTG CCTTATTGGT TTTCTTGCA AACCTTTCTC ATATTTTCTC
751 ATGACCATTG CCAGTGACCA AGGCCCATGT GTGTGTTGTG TGTAAATGTTG
801 GGCATGTACA AGCTTAAATA ACSTGCCGAC AGCAGCTGTT CAAAGTTGGT
851 ATTCATTAGG CTGTTGCCCT CTGGGCTGGA GCTGGCTTAA TCCTGACACC
901 GGCTGCCAGG AGAAAACCTC ATGGATCACA CACCAACCT TAATAACAGC
951 ATCCGTGACC TGCACCTCC AGTACAGAAT CCCAGCCCAA GAAGCTTCTT
1001 ATGTAGTTGT ATATTTAAT GAACTGCTAC GGGGTAGCAG GCACAGAGCT
1051 TCACTTTTGT GCTCTACAGA AACCCCAAGC GGGGTAGCAG GCACAGAGCT
1101 TTGAATAACT CTTTCTAAC ACTAAATCTG CCAACACAGA CAGAGCAGCT
1151 CACACGTATA GGCAGGTGTG AGGACAGAGT GCTAAGAAAT GCCTGCTCCC
1201 TCTGCATGCT CTTTCTTGT TTCCAAGTCC AATCAAGTGA TCCTGGGAAA
1251 CAAATCTGTC TGGATTGCGG AGGGTGGTTC TGAAGAAGCT GCCAAGACGT
1301 TAAAGAAGGG TGAAGAGTAG GCAGAATATA AGTAGTAAC CTGAGTCAAG
1351 ACTCTCAAAA GCTAGCAGCC TGATGACAAT AGGATTTATT TCAGCCAGGA
1401 TAGTGTCTGT CTGTGAGTGC ATCATTTTAA GACAGTATGA CTTCATGTTG
1451 TTACAAACTA TGTATAGTAT GTATGTTTGT TGGGTTGTAT ATATACATAA
1501 TATATATTAT ATATATATAT GAGAGATTG GTGACTTTTG ATACGGGTTT
1551 GGTGCAGGTG AATTTATTAC TGAGCCAAAT GAGGCACATA CCGAGTCAGT
1601 AGTTGAAGTC CAGGGCATTG GATACGTTT ATGATTTCCTA TATATGTATA
1651 GTGCCTATCC CATGCTGTAG TCACTGTTAT GTTAAATCCA GAAGTTACAC
1701 TAGAGCCAGC GATACTTTAT TTGTAGACAA TCAATTTGAA TCCATATGTT
1751 ATTACTGGCA GATGATACAT GATTACAGT CTGAATCTGT AACACTTACA
1801 AAAGGAAACC CAGAGCAGCT TGATGAGTTT TTGTTCTGCT TTCGTTCTCG
1851 GGAGTCAGTA GAAACAGCAG TTGATGTGG TTATGTTAGT CTCAGATAC
1901 TTAATTTGTT GACCTTACTT CAGAAAAATT TTGTATGTAT TATATTTGTG
1951 GGAAGGTAAA ATAATCATT GAGATTTTTA TCAATATGA AGATTAGTTA
2001 TTTATGAAAA ACAAGAAAT GTCTATTTT CTTTGTCCC AATTAATGTA
2051 GATAAATTTT AAAATGCATT AAAGTAATGG TCCGAAAAA AAAAA
```

## BLAST Results

-----

No BLAST result

Medline entries  
-----

89098903:  
Human interleukin 7: molecular cloning and growth factor  
activity on human and murine B-lineage cells.

Peptide information for frame 2  
-----

ORF from 368 bp to 679 bp; peptide length: 104  
Category: similarity to known protein

1 METSHAHESN CKIKGYGVVQ QLLHSQLCGA GEDLGPIGVS YVYSFRAVPF  
51 SLILSNASLH SLGGKDSIQV GCLKELMGPL SELADQILGN LFNFYDFPSH  
101 ILKW

## BLASTP hits

Entry B32223 from database PIR:  
interleukin-7 precursor (clone 1) - human  
Score = 66, P = 7.0e-01, identities = 21/70, positives = 33/70

## Alert BLASTP hits for DKFZphtes3\_35e21, frame 2

PIR:B32223 interleukin-7 precursor (clone 1) - human, N = 1, Score =  
66, P = 0.72

TREMBL:PADAL1\_1 gene: "dall"; P.abies dall mRNA, N = 2, Score = 59, P  
= 0.77

PIR:C32223 interleukin-7 precursor (clone 4) - human, N = 1, Score =  
66, P = 0.79

TREMBL:PRU76726\_1 gene: "PrMADS3"; product: "MADS-box protein"; Pinus  
radiata MADS-box protein (PrMADS3) mRNA, complete cds., N = 2, Score =  
59, P = 0.94

>PIR:B32223 interleukin-7 precursor (clone 1) - human  
Length = 133

## HSPs:

Score = 66 (9.9 bits), Expect = 1.3e+00, P = 7.2e-01  
Identities = 21/68 (30%), Positives = 33/68 (48%)

Query: 39 VSYVYSFRAVPFSLIL-----SNASLHSLGGK--DSIQVGCLKELMGPLSELADQILGNL 91  
VS+ Y F P L+L S+ + GK +S+ + + +L+ + E+ L N  
Sbjct: 4 VSFRIYFGLPPLILVLLPVASSDCDIEGKDGKQYESVLMVSIQQLLDSMKIEGNSCLNNE 63  
  
Query: 92 FNFYDFPSHI 101  
FNF F HI  
Sbjct: 64 FNF--FKRHI 71

Pedant information for DKFZphtes3\_35e21, frame 2  
-----

## Report for DKFZphtes3\_35e21.2

[LENGTH]	104
[MW]	11339.12
[pI]	5.87
[PROSITE]	MYRISTYL 2
[PROSITE]	PKC_PHOSPHO_SITE 1
[PROSITE]	ASN_GLYCOSYLATION 1
[KW]	Alpha_Beta

SEQ METSHAHESNCKIKGYGVVQQLLHSQLCGAGEDLGPIGVS YVYSFRAVPFSLILSNASLH  
PRD ccchhhhhcccccccchhhhhhhhhhhcccccccccccccccccccccccccccccccccccc

WO 01/12659

PCT/IB00/01496

SEQ SLGGKDSIQVGCCLKELMGPLSELADQILGNLFNFYDFPSHILKW  
PRD cccccceccccccccchhhhhhhcccccccccccccc

Prosite for DKFZphtes3\_35e21.2

PS00001	56->60	ASN_GLYCOSYLATION	PDOC00001
PS00005	44->47	PKC_PHOSPHO_SITE	PDOC00005
PS00008	63->69	MYRISTYL	PDOC00008
PS00008	89->95	MYRISTYL	PDOC00008

(No Pfam data available for DKFZphtes3\_35e21.2)

DKFZphtes3\_35g6

group: testes derived

DKFZphtes3\_35g6 encodes a novel 482 amino acid protein with high partial similarity to H. sapiens chromosome 19, cosmid R27216.

No informative BLAST results: No predictive prosite, pfam or SCOP motive.

The new protein can find application in studying the expression profile of testis-specific genes.

strong similarity to R27216\_1

complete cDNA, complete cds, EST hits

Sequenced by DKFZ

Locus: /map="15"

Insert length: 3177 bp

Poly A stretch at pos. 3167, polyadenylation signal at pos. 3148

```
1 GGAGGCAGCG CCGGCCTCCG GAGGCGGCCT GGGCGATGGC GCGGAGTTT
51 TGTCCATAAC CTGGGCAACC GCGCAGCTGG AGGATGGCCT CACTCGGGCC
101 TGCCGAGCTT GGGGAGCAGG CGTCGGGGGC TGAGCGCGAG CCGGGCCCCG
151 CGGGGCCGCC GCCGCCGCC TCACCGTCCT CTCTGGGGCC CCTGCTCCCC
201 CTGACGCGGG AACCTCTCTA CAACTGGCAG GCGACCAAGG CGTCGCTGAA
251 GGAGCGCTTC GCCTTCCTCT TCAACTCGGA GCTGCTGAGC GATGTCGCTC
301 TCGTACTGGG CAAGGGTCGC GCGCGCGCG CGCTGGGGGG CCGCAGCGCC
351 ATCCCCGCCC ACCGCTTCGT GCTGGCGGCC GGCAGCGCGG TCTTTGACGC
401 CATGTTCAAC GCGCGCATGG CCACCACGTC GCGCGAGATC GAGCTGCCGG
451 ACGTGGAGCC CGCAGCCTTC CTGGCGCTGC TGAGATTTCT ATATTGAGAT
501 GAAAGTTCAA TTGGTCCAGA AACAGTTATG ACCACTCTTT ATACTGCCAA
551 GAAATACGCA GTCCAGCCCT TGGAGCACA CTGTGTAGAA TTCTCACCAC
601 AACATCTTAG GGCAGATAAT GCCTTTATGT TACTTACTCA GGCTCGATTA
651 TTTGATGAAC CTCAGCTTGC TAGTCTTTGT CTAGATACAA TAGACAAAAG
701 CACAATGGAT GCAATAAGTG CAGAAGGGTT TACTGATATT GATATAGATA
751 CACTCTGTGC AGTTTATAG AGAGACACAC TCAGTATTCG AGAAAGTCGA
801 CTTTTTGGAG CTGTTGTACG CTGGGCAGAA GCAGAAATGC AGAGACAACA
851 ATTACCTGTG ACTTTTGGGA ATAAACAAAA AGTTCTAGGA AAGSCACTTT
901 CCTTAATCCG GTTCCCACTG ATGACAATTG AGGAATTTGC AGCAGGCTCT
951 GCTTAATCTG GAATTTGTGC AGATCTGAA CTGGTAAACC TCTTCTCTCA
1001 TTTTACTCTC AACCTTAAC CCCGAGTTGA ATCATTGAC CGACCAAGAT
1051 GCTGTCTCAG GGGAAAGGAA TSGTCATCA ATAGATTCCA CAAAGTAGAA
1101 AGCCCGCTGG GTTACAGTGG GACGAGTGAT CGAATCAGAT TCACAGTTAA
1151 TAGAAGGATC TCTATAGTTG GATTGGCTTT GTATGGATCT ATTCTAGGCC
1201 TCACAGATTA TCAAGTGAAT ATACAGATCA TTGAATATGA GAAAAGACAA
1251 ACCCTGGGAC AGAATGATAC CGGCTTTAGT TGTGATGGGA CAGCTAACAC
1301 ATTCAGGGTC ATGTTCAAGG AACCATAGA GATCCTGCCC AATGTGTGCT
1351 ACACAGCATG TGCAACACTC AAAGGTCCAG ATTCCCACTA TGGCACAATA
1401 GGATTGAAGA AAGTAGTGCA TGAGACACCT GCTGCAAGCA AGACTGTTTT
1451 TTTCTTTTTT AGTTCCCTG GCAATAATAA TGGCACTTCA ATAGAAGATG
1501 GACAAATCCC AGAAATCATA TTTTATACAT AATTTAGCAT TATAATACAT
1551 CTTGGCTAAA TAATACCATA CAATCTAGTG TCAAAAACAT AAATGGCCAC
1601 AAAAAAGTAG TTTAGTGTT ATGAATATTT AAAATTGTAA GATAAGAAAC
1651 AGTTTCTTAG AGCAGATAGA AAAATGCTTA TTTAAATCTT TGCAATGATT
1701 AAAAAACAGT TTTCCATTTT CTTACAACCT TAAGAGAAAA GAACTGGGTT
1751 TAATGGTTTA AAAAAAGCA CAGCTTTTTC ACCTTCATCT TGTATAATTT
1801 CATAGATTGG CTGACTAGG GTCTTTCAAT AGTTTGGGAA TTGAAAGATT
1851 CTTGTTATAT ATAGCTAGTT TGGGTTTGT TTTGTTTAA CTATTTTGAA
1901 GGTTAGGTGA GATGGGCAAA TAGGCTTAAC TATTTTGAA GTTGGATGAA
1951 AAGAGATGGG TCAGTATTC TACAGAATTC TTATTAACCT AAATAACTAA
2001 ATTCAGAAAA ATTAAGAAGC TGACTTTATA TTTGGTGGTT TGAAGTATCT
2051 TGTGTTAGC ATTTGTAATA ATGCTAAAAA AGGCCTAATA AAATGCCCAA
2101 GAAATATTC AGTGCAATTA TAGAGAAGGA TATTTGTAG TAGTATAGTA
2151 ATGTGTTATG TAGTACAGTT TTAAGCTAT AAATGGAATT TTGTGTAAT
2201 TCACAAAAAT GTGATATAAA CAGGATCTAA GACTGGATTC CCTGTCACTA
2251 AACTGCACCA CTATACCTGT CTCTCTGTGT GGGGGACACT GCTGATGATT
2301 CCCAAGATTG AGATGATGAC GGTGATGACG ACTGGGTGAA CAGCCATCAC
2351 TTCAACATTG TGATAATCCT TCACAGCAAG AACCCGAATA AAATACTAAC
2401 ATTTCTAACA ACTGCTCTGA CATTGTAAG AGATCCAACA GAATCACTCC
2451 TGCTGAAAAA TACGCTTCTT GCCACCTACA CATTCTATT TAGGAAAGTA
2501 AATTGCTTC ATGGTCATGA CCCCATTAGT CAGTGTTACA GCTGTGTTGG
2551 GGATAGGAAG TATATCTGGC AGATTGACAT TTATACACTT TTTTATAAAG
2601 CAGATTTTAA AATATAGTAA CATCCATTTT TTTCCCTTGA AAGTGATTCT
2651 CTTATAAAAA ATGAAAGTGG AGTTTAAGGT ATATCAAAAT GTTGGAAG
2701 GTGATAAAAA ATCAAAATTC TTTTAAATAT CAACTTAATT TTTTCTAAGT
```

```
2751 AAGATACAAA AAATTTTCAT CTAAAGTAAT ATTTCACTTT ATATTGTAAA
2801 GAAGGTAGGT ATATTGGTGG CTGAGGTCTC TTGAAATTGC TAAAGGGAAA
2851 TTTTCTATG GTAATGCTCT TACGGATATA AGCCTCAGTT AAATGGAAAT
2901 ATCTATGGGA TGTGTGGTTC TGGTTAACTA AAAATTAACC AGTAAACACT
2951 CTGTAGTAAC CATTACAGAA AATACTTCTG CCTTAAAAAA TATGATATGC
3001 CAGAGATGAG TTAGTGTTC TTAGCGTTGG AGACCTATAA ATGCCTCATC
3051 TGTGTACTG AACAAATTGAA ACTGCATGCA GCCATAAAG GGACAAGAAA
3101 CAGAACTGTT TACTAACTTT GGGACATCCC CTGGAGTTT TAAAAATAAA
3151 TAAATATATA TATATATAAA AAAAAA
```

## BLAST Results

Entry G37753 from database EMBL:  
SHGC-63477 Human Homo sapiens STS genomic.  
Score = 1627, P = 3.0e-66, identities = 327/329

Entry G37752 from database EMBL:  
SHGC-63476 Human Homo sapiens STS genomic.  
Score = 1578, P = 6.2e-64, identities = 320/324

## Medline entries

No Medline entry

## Peptide information for frame 3

ORF from 84 bp to 1529 bp; peptide length: 482  
Category: similarity to unknown protein

```
1 MASLGPAAG EQASGAEEF GPAGPPPPPS PSSLGPLLPL QREPLYNWQA
51 TKASLKERFA FLFNSLELSD VRFVLGKGRG AAAAGGPQRI PAHRFVLAAG
101 SAVFDAMFNG GMATTSAEIE LPDVEPAFL ALLRFLYSDE VQIGPETVMT
151 TLYTAKKYAV PALEAHCVEF LTKHLRADNA FMLLTQARLF DEPQLASLCL
201 DTIDKSTMDA ISAEFTDID IOTLCAVLER DTLSIRESRL FGAVVRWAEA
251 ECQRQQLPVT FGNKQKVLGK ALSLIRFPLM TIEEFAAGPA QSCILSDREV
301 VNLFLHFTVN PKPRVEYIDR PRCLLRGKEC CINRFQOVES RWCYSGTSDR
351 IRFTVNRRIIS IVGFLYGS I HGPTDYQVNI QIIIEVEKKQT LGQNDTGFSC
401 DGTANTFRVM FKEPIELPN VCYTACATLK GPDSHYGKTK LKKVVHETPA
451 ASKTVEFFFS SPGNNGTST EDGQIPEIIF YT
```

## BLASTP hits

Entry AC005306.2 from database TREMBL:  
product: "R272I6.1"; Homo sapiens chromosome 19, cosmid R27216,  
complete sequence.  
Score = 1298, P = 1.9e-132, identities = 245/297, positives = 268/297

Entry CEF38H4.9 from database TREMBLNEW:  
gene: "F38H4.7"; Caenorhabditis elegans cosmid F38H4  
Score = 1237, P = 5.6e-126, identities = 248/446, positives = 322/446

Entry AC004678.1 from database TREMBL:  
product: "R34094.1"; Homo sapiens chromosome 19, cosmid R34094,  
complete sequence.  
Score = 555, P = 1.0e-53, identities = 112/137, positives = 123/137

Alert BLASTP hits for DKFZphtes3\_35g6, frame 3

No Alert BLASTP hits found

Pedant information for DKFZphtes3\_35g6, frame 3

Report for DKFZphtes3\_35g6.3

[LENGTH]	482
[MW]	52771.47
[pI]	5.79

```

(HOMOL) TREMBL:AC005306.2 product: "R27216_1"; Homo sapiens chromosome 19, cosmid
R27216, complete sequence. 1e-142
[BLOCKS] BL01075D Acetate and butyrate kinases family proteins
[SUPFAM] POZ domain homology 3e-08
[SUPFAM] A55R protein middle region homology 5e-06
[SUPFAM] A55R protein 5e-06
[SUPFAM] A55R protein carboxyl-terminal homology 5e-06
[PROSITE] MYRISTYL 6
[PROSITE] CAMP_PHOSPHO_SITE 2
[PROSITE] CK2_PHOSPHO_SITE 9
[PROSITE] TYR_PHOSPHO_SITE 1
[PROSITE] PKC_PHOSPHO_SITE 7
[PROSITE] ASN_GLYCOSYLATION 2
[KW] Alpha_Beta
[KW] LOW_COMPLEXITY 11.20 %

```

```

SEQ MASLGPAAGGEQASGAEEAGPGAGPPPPSPSSLGPLLPLOREPLYNQATKASLKERFA
SEG ...XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
PRD cccccccchhhhhhhccccccccccccccccccccccccccccccccccccchhhhhhhhhhhhh

SEQ FLFNSSELLSDVRFLGKGRGAAAGGPQRI PAHREVLAAAGSAVDFAMFNGMATTSAEIE
SEG .....XXXXXXXXXXXX
PRD hhhccccccccccccccccccccccccccccchhhhhheccccchhhhhhhcchhhhhhhhe

SEQ LPDVEPAALFALLRFLYSDEVQIGPETVMTTLTYAKKYAVPALEAHCVFELTKHLRADNA
SEG .....
PRD eccccchhhhhhhhhhhccccchhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhccch

SEQ FMLLTQARLFDEPQLASLCLDTIDKSTMDAISAEFTDIDITLCAVLEROTLSIRESRL
SEG .....
PRD hhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhccccchhhhhhhhh

SEQ FGAVVRWAEAEQRRQLPVTFGNKQKVLGKALSIRFPLMTIEEFAAGPAQSGILSDREV
SEG .....
PRD hhhhhhhhhhhhhhhhhhhhhhhccccchhhhhhhhhhhhhhhheccccccccccccchhhhh

SEQ VNLFLHFTVNPKRVEYIDRPRCLRGKECCINRFQVESRWGYSGTSDRIFTVNRRIIS
SEG .....
PRD hhhhhheccccccccccccccccccccccccchhhhhhhhhhhccccccccccccchhhhhheee

SEQ IVGFGLYGSINGPTDYQVNIQIIIEYEKKQTLGQNDTGFSKCDGTANTFRVMFKEPIELPN
SEG .....
PRD eccccccccccccchhhhhhhcchhhhhhhhhcccccccccccccccccccccccccccc

SEQ VCYTACATLKGPDHYGTGKLKKVVHETPAASKTVFFFSPPGNNGTSIEDGOIPEIIF
SEG .....
PRD ccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc

SEQ YT
SEG ..
PRD cc

```

## Prosites for DKFZphtes3\_35g6.3

PS00001	394->398	ASN_GLYCOSYLATION	PDOC00001
PS00001	466->470	ASN_GLYCOSYLATION	PDOC00001
PS00004	357->361	CAMP_PHOSPHO_SITE	PDOC00004
PS00004	387->391	CAMP_PHOSPHO_SITE	PDOC00004
PS00005	54->57	PKC_PHOSPHO_SITE	PDOC00005
PS00005	154->157	PKC_PHOSPHO_SITE	PDOC00005
PS00005	234->237	PKC_PHOSPHO_SITE	PDOC00005
PS00005	296->299	PKC_PHOSPHO_SITE	PDOC00005
PS00005	348->351	PKC_PHOSPHO_SITE	PDOC00005
PS00005	406->409	PKC_PHOSPHO_SITE	PDOC00005
PS00005	428->431	PKC_PHOSPHO_SITE	PDOC00005
PS00006	14->18	CK2_PHOSPHO_SITE	PDOC00006
PS00006	54->58	CK2_PHOSPHO_SITE	PDOC00006
PS00006	115->119	CK2_PHOSPHO_SITE	PDOC00006
PS00006	206->210	CK2_PHOSPHO_SITE	PDOC00006
PS00006	217->221	CK2_PHOSPHO_SITE	PDOC00006
PS00006	234->238	CK2_PHOSPHO_SITE	PDOC00006
PS00006	281->285	CK2_PHOSPHO_SITE	PDOC00006
PS00006	296->300	CK2_PHOSPHO_SITE	PDOC00006
PS00006	468->472	CK2_PHOSPHO_SITE	PDOC00006
PS00007	430->437	TYR_PHOSPHO_SITE	PDOC00007
PS00008	80->86	MYRISTYL	PDOC00008
PS00008	110->116	MYRISTYL	PDOC00008
PS00008	365->371	MYRISTYL	PDOC00008

WO 01/12659

PCT/IB00/01496

PS00008	392->398	MYRISTYL	PDOC00008
PS00008	402->408	MYRISTYL	PDOC00008
PS00008	463->469	MYRISTYL	PDOC00008

(No Pfam data available for DKFZphtes3\_35g6.3)



DKFZphtes3\_35k16  
-----

group: metabolism

DKFZphtes3\_35k16 encodes a novel 666 amino acid protein with weak similarity to fatty acid-CoA synthetaseses/ligases.

The novel protein contains a putative AMP-binding domain signature, which is present in enzymes, which act via an ATP-dependent covalent binding of AMP to their substrate. This domain is found in several CoA synthetases, such as acetate-CoA ligase (EC 6.2.1.1), long-chain-fatty-acid-CoA ligase (EC 6.2.1.3), bile acid-CoA ligase. Therefore it is a new fatty acid-CoA synthetaseses/ligase with unknown substrate.

The new protein can find application in modulation of fatty acid metabolism and as a new enzyme for biotechnologic production processes.

similarity to acyl-CoA synthetase

complete cDNA, complete cds, potential start codon at Bp 50,  
few EST hits, seems to be a testis specific cDNA,  
5 of 6 EST hits are from testis derived libraries

Sequenced by DKFZ

Locus: unknown

Insert length: 2520 bp

Poly A stretch at pos. 2510, polyadenylation signal at pos. 2490

```
1 CAGATGTCCC AGCTCCAGTG CTGTGGAGCA TGGTTTCTGC ACACCTGGAA
51 TGACTGGAAC CCCAAGACT CAAGAAGGAG CTAAAGATCT TGAAGTAGAC
101 ATGAATAAAA CAGAAGTTAC TCCCAGGCTG TGGACCACCT GTCGAGATGG
151 AGAAGTCCTT CTGAGGCTAT CCAACACGG ACCAGGCCAT GAGACCCCGA
201 TGACCATCCC TGAATTTTTT CGAGAGTCAG TCAACCGATT TGAACCTTAT
251 CCAGCCCTCG CATCCAAGAA TGGCAAAAAG TGGGAAATTC TGAATTTCAA
301 CCAGTACTAT GAGGCTCTC GGAAGCTGC AAAATCCTTG ATCAAGCTGG
351 GTTGTGAGCG TTTCCACGGA GTTGCTATCC TGGGGTTTAA CTCTGCACAG
401 TGGTTATACA CTGCTGTGGT TGCCATCCTA GCCGGGGGTC TTTGTGTTGG
451 TATTTATGCC ACCAACTCTG CCGAGGCTTG TCAATATGTC ATCACTCATG
501 CCAAAGTGAA CATCTTGCTG GTTGAGAAATG ATCAACAGTT ACAGAAAAATC
551 CTTTCGATTC CACAGAGCAG CTTAGAGCCC CTAAAAGCGA TCATCCAGTA
601 CAGACTGCCA ATGAAGAAGA ACAACAACCT GTACTCTTGG GATGATTTCa
651 TGGAACTTGG CAGAAGTATC CCTGACACCC AACTGGAGCA GGTCACTCGAG
701 AGCCATAAGG CGAATCAATG CGCAGTGCTC ATCTACACTT CAGGGACCAC
751 AGGCATACCC AAGGGAGTGA TGCTCAGTCA TGACAACATC ACGTGGATTG
801 CAGGAGCAGT GACAAAGGAC TTTAAACTGA CAGACAAGCA TGAGACGGTG
851 GTTAGCTACC TCCCACTCAG CCATATTGCA GCACAGATGA TGGACATCTG
901 GGTACCCATA ARGATTGGGG CGCTCACATA CTTTGCTCAA GCAGATGCTC
951 TCAAGGGCAC CTGGTAAGT ACTCTAAAGG AGGTAAAACC TACTGTCTTC
1001 ATTGGAGTGC CTCAAATTTG GGAGAAGATA CATGAGATGG TGAAGAAAAA
1051 TAGTGCCCAAG TCCATGGGCT TGAAGAGAA GGCATCTCTG TGGGCAAGAA
1101 ACATTGGCTT CAAGGTCAAC TCAAAAAGA TGTGGGGAA ATATAATACT
1151 CCGTGAGCT ACCGCATGGC TAAGACTCTC GTGTTCAGCA AAGTCAAGAC
1201 ATCCCTTGGC TTGGATCACT GTCACTCTTT TATCAGTGGG ACTGCGCCCC
1251 TCAACCAAGA GACTGCCGAG TTCTTTCTAA GCTTGGACAT ACCTATAGGC
1301 GAGTTGTATG GGTGAGTGA GAGCTCGGGA CCCCACACGA TATCCAACCA
1351 GAATAACTAC AGGCTTCTAA GCTGTGGCAA GATCTTGACT GGGGTGAAGA
1401 ATATGCTGTT CCAGCAGAAC AAGGATGGCA TTGGGGAGAT CTGCCTCTGG
1451 GGTAGGCACA TCTTCATGGG CTATCTGGAA AGTGAGACTG AACTACAGA
1501 GGCCATCGAT GATGAAGGCT GGCTACACTC TGGGGATCTG GGCCAGCTGG
1551 ACGGTCTGGG TTTCTCTAT GTCAACGGCC ACATCAAGA AATCCTTATC
1601 ACTGCTGGTG GTGAAATGT GCCCCCCATT CCTGTTGAGA CTTTGGTTAA
1651 GAAGAAGATC CCCATCATCA GTAACGCCAT GTTAGTAGGA GATAAATGA
1701 AGTTTCTGAG CATGTTGCTG ACCTGAAGT GTGAGATGAA TCAGATGAGC
1751 GGAGAACCCT TGGACAAGCT GAACCTCGAG GCCATCAACT TCTGTGGGGG
1801 TCTGGGCAGC CAGGCATCCA CCCTGACTGA GATGGTGAAG CAGCAAGACC
1851 CCCTGGTCTA CAAGGCCATC CAGCAAGCCA TCAATGCTGT GACCAAGAA
1901 GGCATGAACA ATGCACAGAG GATTGAAAAG TGGGTCACTT TGGAGAAGGA
1951 CTTTTCATC TATCGTGGAG AGCTAGGTCC AATGATGAAA CTTAGAGAC
2001 ATTTTGTAGC CCAGAAATAC AAAAAACAAA TTGATCACAT GTACCACTGA
2051 CTGCTTTGAT GGAGCTGCTC TCAGCTGTTT TGATGCCTTC AGCAGGAAGA
2101 CCTCATTTGA ATAAGTGAAA TGCTGTCTTA GGTAGAAGCT CTCCTGTCTG
2151 TTTTAAAGAA GCCACATTCC TCATTGGTCA GTTCTTGAT TGTTCGTCTG
2201 TTGGAGAGGT GCTCCCTAGA AGAACCTGCC ATACGTTTCA AAGCAATAAA
2251 ATCACTGTAT ATCTTTCTAA GGACCTTCAA GTCATGACTC CAGGGAAGCC
2301 TATTGGGAAG TCTACTAAAA ACTGCCTGAT TTACAAGAAA GACCTGAAC
```

2351 TGTGGGCTCC CATTGATTT TTTTCTCTC AGGGGACTCA GACATTAGAA  
2401 AGAAAAAGCC TCACAGATTT GAAGAACTGG ACCCCCAAAAT CAACTCACCT  
2451 GCCTGGAAGC AACTGGGAAA CCCTTCCAAT AAGTCTGAT AATAAAGCAC  
2501 TTCAGGGTCC AAAAAAAAAA

## BLAST Results

No BLAST result

## Medline entries

No Medline entry

## Peptide information for frame 2

ORF from 50 bp to 2047 bp; peptide length: 666  
Category: similarity to known protein

1 MTCTPKTQEG AKOLEVDMNK TEVTPRLWTT CRDGEVLLRL SKHGPGHETP  
51 MTIPEFFRES VNRFGTYPAL ASKNGKKWEI LNFNOYYEAC RKAASLILK  
101 GLERFHGCVI LGFNSAEWFI TAVGAILAGG LCVGIYATNS AEACQYVITH  
151 AKVNILLVEN DQLOKILSI POSSLEPLKA IIQYRLPMKK NNNLYSWDDF  
201 MELGRSIPDT OLEOVIESOK ANQCAVLIYT SGTGIPKGV MLSDNITWI  
251 AGAVTKDFKL TDKHETVVS YRLSHIAAQM MDIWPVPIKIG ALTYFAQADA  
301 LKGLTVSTLK EVKPTVFIGV PQIWEKIHQM VKKNSAKSMG LKKKAFVWAR  
351 NIGFKVNSKK MLGKYNTFVS YRMAKTLVFS KVKTSGLLDH CHSFISGTAP  
401 LNQETAFFEL SLDIPIGELY GLESSSGPHT ISNQNRYRL SCGKILTCK  
451 NMLFQONKDG IGEICLWGRH IFMGYLESET ETTEAIDDEG WLHSGDLGQL  
501 DGLGFLYVTG HIKEILITAG GENVPPPIVE TLVKKKIPPI SNAMLVGDKL  
551 KFLSMLTLK CEMNQMSGEP LDKLNFEAIN FCRGLGSOAS TVTEMVKKQD  
601 PLVYKAIQGG INAVNQEAMN NAQRIEKWI LEKDFSIIYG ELGPMMLKLR  
651 HFVAQRYKKQ IDHMYH

## BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphtes3\_35k16, frame 2

TREMBL:AB014531.1 gene: "KIAA0631"; product: "KIAA0631 protein"; Homo  
sapiens mRNA for KIAA0631 protein, partial cds., N = 1, Score = 1641, P  
= 8.9e-169

PIR:E70937 probable fadD15 - Mycobacterium tuberculosis (strain H37RV),  
N = 2, Score = 532, P = 3.6e-62

PIR:H64041 long-chain-fatty-acid--CoA ligase homolog - Haemophilus  
influenzae (strain Rd KW20), N = 2, Score = 486, P = 6.5e-59

>TREMBL:AB014531.1 gene: "KIAA0631"; product: "KIAA0631 protein"; Homo  
sapiens mRNA for KIAA0631 protein, partial cds.  
Length = 634

## HSPs:

Score = 1641 (246.2 bits), Expect = 8.9e-169, P = 8.9e-169  
Identities = 319/628 (50%), Positives = 440/628 (70%)

Query: 38 LRLSKHGPGHETPMTIPEFFRESVNRFGTYPALASKNGKKWEILNFNOYYEACRKAASL 97  
LR+ P + P T+ F E++++G AL K RWE +++QY R+AAK  
Sbjct: 2 LRIDPSCP--QLPYTVHRMFYEALDKYGDIALGFKRQDKWEHISYSQYLLARRAAGF 59  
Query: 98 IKLGLERFHGCVILGFNSAEWFI TAVGAILAGGLCVGIYATNSAEACQYVITHAKVNILL 157  
+KLG++ H V ILGFNS EWF +AVG + AGG+ GY T+S EACQY+ N+++  
Sbjct: 60 LKGLKQAHSAVILGFNSPEWFFSAVGTVFAGGIVTGIYTTSSPEACQYIAYDCCANVM 119  
Query: 158 VENDORLOKILSIPOSSLEPLKAIQYRLPM-KNNNLYSWDDFMELGRSIPDTOLEQVI 216  
V+ +QL+KIL I L LKA++ Y+ P K N+Y+ ++FMELG +P+ L+ +I  
Sbjct: 120 VDTQKQLEKILKI-WKQLPHLKAVVIYKEPPPNKMANVYTMEEFMELGNEVPEEALDAII 178  
Query: 217 ESQANQCAVLIYTS GTTGIPKGVMLSDNITWIA--GAVTKDFKLT-DKHETVVSYLPL 273

```

++Q+ NQC VL+YTS GTTG PKGVMLS DNITW A G+ D + + + E VVSYLPL
Sbjct: 179 DTQQPNQCCVLVYTS GTTGPNKGVMLSQDNITWTARYGSQAGDIRPAEVQQEVVSYLPL 238
Query: 274 SHIAAQMDIWPVPIKIGALTYFAQADALGKTLVSTLKEVKPTVFIVGPQIWEKIHVMVK 333
SHIAAQ+ D+W I+ GA FA+ DALKG+LV+TL+EV+PT +GVP++WEKI E +++
Sbjct: 239 SHIAAQIYDLWTGIQWGAQVCFAEPDALKGLSVNTLREVEPTSHMGVPRVWEKIMERIQE 298
Query: 334 NSAKSMGLKKKAFVWNRNIGFKVNSKKMLGKYNTPVSYMAKTLVFSKVKTSGLDHCSS 393
+A+S ++K +WA ++ + N G P + R+A LV +KV+ +LG C
Sbjct: 299 VAAQSGFIRKMLLMWMSVTLEQNLTPGSDLKPFTRLDYLVLAQRQALGFACQK 357
Query: 394 FISGTAPLNQETAFFFLSLDIPIGELYGLSESSGPHITISQNNYRLSCGKILTCKNML 453
G AP+ ET FFL L+I + YGLSE+SGPH +S+ NYRL S GK++ GC+ L
Sbjct: 358 NFYGAAPMAETQHFLLNIRLYAGYGLSETSGPHMSSPYNYRLYSSGKLVPGCRVKL 417
Query: 454 FOQNKDGI GEICLWGRHIFMGYLESETETETEAIDDEGLHSGDLGLDGLFLYVTGHIK 513
Q+ +GIGEICLWGR IFMGYL E +T EAID+EGWLH+GD G+LD GFLY+TG +K
Sbjct: 418 VNQDAEGIGEICLWGRITIFMGYLNMDKTCEAIDDEGLHSGDAGRLDADGFLYITGRK 477
Query: 514 EILITAGGENVPPIPVETLVKKKIPISNAMLVGDGLKFLSMLLTLCENMQMSGEPLDK 573
E++ITAGGENVPP+P+E VK ++PIISNAML+GD+ KFLSMLLTLC ++ + + D
Sbjct: 478 ELIITAGGENVPPVPIEEAVKMLPIISNAMLIGDQKFLSMLLTLCPLDPTSDQTDN 537
Query: 574 LNFEAINFCRGLGSQASTVTEMVKQDPLVYKAIQGINAVNQEAMNNAQRIEKWVILEK 633
L +A+ FC+ +GS+A+TV+E++++D VY+AI++GI VN A++ I+KW ILE+
Sbjct: 538 LTEQAVEFCQVRGSRATTYSEIEKKDEAVYQATEEIGIRRVNMNAAAPYHIQKWAILER 597
Query: 634 DFSIYGGEI GPMMLKRRHFVAQKYKKQIDHMY 665
DFSI GGELGP MKLKR V +KYK ID Y
Sbjct: 598 DFSISGGEI GPTMKLRLTVLEKYKGIIDSFY 629

```

Pedant information for DKFZphtes3\_35k16, frame 2

#### Report for DKFZphtes3\_35k16.2

```

[LENGTH] 666
[MW] 74344.97
[PI] 8.67
[ROMOL] TREMBL:AB014531.1 gene: "KIAA0631"; product: "KIAA0631 protein"; Homo sapiens
mRNA for KIAA0631 protein, partial cds. 1e-176
[FUNCAT] 1 lipid metabolism [H. influenzae, HI0002] 2e-55
[FUNCAT] 08.10 peroxisomal transport [S. cerevisiae, YER015w] 2e-29
[FUNCAT] 30.19 peroxisomal organization [S. cerevisiae, YER015w] 2e-29
[FUNCAT] 01.06.13 lipid and fatty-acid transport [S. cerevisiae, YER015w] 2e-29
[FUNCAT] 01.06.07 lipid, fatty-acid and sterol utilization [S. cerevisiae, YER015w]
2e-29
[FUNCAT] 01.06.01 lipid, fatty-acid and sterol biosynthesis [S. cerevisiae, YMR246w]
2e-23
[FUNCAT] 06.07 protein modification (glycosylation, acylation, myristylation,
palmitoylation, farnesylation and processing) [S. cerevisiae, YMR246w] 2e-23
[BLOCKS] BL00455
[SCOP] d1lci_ 5.19.1.1.1 Luciferase [Firefly (Photinus pyralis)] 1e-49
[EC] 1.13.12.7 Photinus-luciferin 4-monooxygenase (ATP-hydrolysing) 9e-17
[EC] 6.2.1.3 Long-chain-fatty-acid--CoA ligase 4e-34
[EC] 5.1.1.11 Phenylalanine racemase (ATP-hydrolysing) 6e-08
[EC] 6.2.1.12 4-Coumarate--CoA ligase 8e-18
[PIRKW] duplication 6e-07
[PIRKW] phosphopantetheine 3e-12
[PIRKW] multifunctional enzyme 3e-06
[PIRKW] ligase 6e-08
[PIRKW] acid-thiol ligase 4e-34
[PIRKW] transmembrane protein 5e-22
[PIRKW] monooxygenase 9e-17
[PIRKW] hydrolase 4e-34
[PIRKW] peroxisome 9e-15
[PIRKW] antibiotic biosynthesis 3e-12
[PIRKW] isomerase 6e-08
[PIRKW] flavonoid biosynthesis 1e-17
[PIRKW] magnesium 9e-15
[PIRKW] ATP 5e-22
[PIRKW] oxidoreductase 9e-17
[PIRKW] liver 2e-31
[SUPFAM] alpha-aminoadipyl-cysteine-valine synthetase 3e-07
[SUPFAM] human long-chain-fatty-acid--CoA ligase 4e-34
[SUPFAM] gramicidin S synthetase I 6e-08
[SUPFAM] peptide synthetase ppsE 7e-06
[SUPFAM] gramicidin S synthetase I repeat homology 3e-12
[SUPFAM] peptide synthetase ppsD 2e-07

```

```

[SUPFAM]      probable acyl-CoA ligase medium chain 2e-09
[SUPFAM]      acetate--CoA ligase 8e-10
[SUPFAM]      acetate--CoA ligase homology 4e-54
[SUPFAM]      surfactin synthetase 3e-12
[SUPFAM]      4-coumarate--CoA ligase 8e-18
[SUPFAM]      short-chain alcohol dehydrogenase homology 8e-07
[SUPFAM]      acyl carrier protein homology 2e-29
[PROSITE]     MYRISTYL 12
[PROSITE]     AMP_BINDING 1
[PROSITE]     AMIDATION 1
[PROSITE]     CAMP_PHOSPHO_SITE 1
[PROSITE]     CK2_PHOSPHO_SITE 9
[PROSITE]     TYR_PHOSPHO_SITE 3
[PROSITE]     PKC_PHOSPHO_SITE 10
[PROSITE]     ASN_GLYCOSYLATION 2
[PFAM]        AMP-binding enzymes
[KW]          Irregular
[KW]          3D
[KW]          LOW_COMPLEXITY 1.80 %

SEQ  MTGTPKTOEGAKDLEVDNMNKTEVTPRLWTTTCRDGEVLLRLSKHGPGHETPMTIPEFFRES
SEG  .....
11ci- .....

SEQ  VNRFGTYPALASKNGKKWEILNFNQYAEACRKAASLIKGLERFHVGVILGFNSAEWFI
SEG  .....
11ci- .....

SEQ  TAVGATLAGGLCVGIATNSAEACQYVITHAKVNILLVENDQQLKILSIPOSSLEPLKA
SEG  .....
11ci- .....

SEQ  IIQYRLPMKKNNNLYSWDDFMELGRSIPDTQLEQVIESQANQCAVLIYTSGTTGIPKGV
SEG  .....
11ci- .....

SEQ  MLSDHNTIAGAVTKDFKLTDKHETVVSYLPLSHIAQMMDIWVPIKIGALTYFAQADA
SEG  .....
11ci- .....

SEQ  LKGTIVSTLKEVKPTVFIGVPOIWEKIHVMVKNKSAKSMGLKKKAFVWARNIGFKVNSKK
SEG  .....
11ci- .....

SEQ  MLGKYNTPVSYRMATLVFSKVKTSGLDHCFSISGTAPLNQETAFFLSLDIPIGELY
SEG  .....
11ci- ..... TTTTCEETTTTCCHHHHHHHHHHCCCCBCEE

SEQ  GLSESSGPHTISNQNNYRLSCGKILTGCKNMLFOQNKDGIICLWGRHIFMGYLESET
SEG  .....
11ci- ..... ECGGGTTEEECCCCCEEEETTTEEEETTTCETTEEEETTTCCEETTTTH

SEQ  ETTEAIDDEGWLHSGDLGQLOGLGFLVYTGHIKEILITAGGENVPPIPVETLVKKKIPII
SEG  .....
11ci- ..... HHHHHBTTTCEEEEEETTTEEE-----ECEEETEECHHHHHHHHHT-TTE

SEQ  SNAMLVGDKLKFLSMLLTLCENMNSGEPLDKLNFEAINFCRLGSOASTVTENVKQOD
SEG  .....
11ci- ..... EEEEEEE

SEQ  PLVYKAIQGINAVNQEAMNNAQRIEKWVILEKDFSIYGGELGPMMLKRHFVAQYKKQ
SEG  .....
11ci- .....

SEQ  IDHMYH
SEG  .....
11ci- .....

```

## Prosites for DKFZphtes3\_35k16.2

PS00001	19->23	ASN_GLYCOSYLATION	PDOC00001
PS00001	246->250	ASN_GLYCOSYLATION	PDOC00001
PS00004	332->336	CAMP_PHOSPHO_SITE	PDOC00004
PS00005	4->7	PKC_PHOSPHO_SITE	PDOC00005
PS00005	24->27	PKC_PHOSPHO_SITE	PDOC00005
PS00005	30->33	PKC_PHOSPHO_SITE	PDOC00005
PS00005	218->221	PKC_PHOSPHO_SITE	PDOC00005
PS00005	261->264	PKC_PHOSPHO_SITE	PDOC00005

PS00005	308->311	PKC_PHOSPHO_SITE	PDOC00005
PS00005	335->338	PKC_PHOSPHO_SITE	PDOC00005
PS00005	358->361	PKC_PHOSPHO_SITE	PDOC00005
PS00005	370->373	PKC_PHOSPHO_SITE	PDOC00005
PS00005	558->561	PKC_PHOSPHO_SITE	PDOC00005
PS00006	30->34	CK2_PHOSPHO_SITE	PDOC00006
PS00006	52->56	CK2_PHOSPHO_SITE	PDOC00006
PS00006	173->177	CK2_PHOSPHO_SITE	PDOC00006
PS00006	196->200	CK2_PHOSPHO_SITE	PDOC00006
PS00006	206->210	CK2_PHOSPHO_SITE	PDOC00006
PS00006	210->214	CK2_PHOSPHO_SITE	PDOC00006
PS00006	308->312	CK2_PHOSPHO_SITE	PDOC00006
PS00006	478->482	CK2_PHOSPHO_SITE	PDOC00006
PS00006	591->595	CK2_PHOSPHO_SITE	PDOC00006
PS00007	659->666	TYR_PHOSPHO_SITE	PDOC00007
PS00007	659->666	TYR_PHOSPHO_SITE	PDOC00007
PS00007	597->605	TYR_PHOSPHO_SITE	PDOC00007
PS00008	3->9	MYRISTYL	PDOC00008
PS00008	65->71	MYRISTYL	PDOC00008
PS00008	124->130	MYRISTYL	PDOC00008
PS00008	130->136	MYRISTYL	PDOC00008
PS00008	134->140	MYRISTYL	PDOC00008
PS00008	235->241	MYRISTYL	PDOC00008
PS00008	239->245	MYRISTYL	PDOC00008
PS00008	303->309	MYRISTYL	PDOC00008
PS00008	387->393	MYRISTYL	PDOC00008
PS00008	421->427	MYRISTYL	PDOC00008
PS00008	498->504	MYRISTYL	PDOC00008
PS00008	586->592	MYRISTYL	PDOC00008
PS00009	74->78	AMIDATION	PDOC00009
PS00455	227->239	AMP_BINDING	PDOC00427

## Pfam for DKF2phtes3\_35kl6.2

HMM_NAME	AMP-binding enzymes	
HMM	*TYRELNERANRLARHLRsekGlrPGDiVgIMMDRSMWMIvAnLGIWKAG	
Query	82 NFNQYYEACRKAASLI-KLGLERFHVGILGFNSAEWFITAVGAILAG	129
HMM	GAYVPIDPeYpDERIqYMLSDSGArLLITQrh...HmqRIPdemwvvdH	
Query	130 GLCVGIYATNSAEACQYVITHAKVNILLVENDQQLQRILSIPOSSLEPLK	179
HMM	IivdWe.....WddlWWHedeeNpqpWvdPeDLAYIIY	
Query	180 AIIQYRLPMKNNNLYSWDDFMELGRSIPDTQLEQVIESQKANQCAVLIY	229
HMM	TSGTTGPKGVMIHrNivNycqWMnWRYGMeedDRILWftSDpYWFda	
Query	230 TSGTTGIPKGVMLSHDNITWAGAVTKDFKLTDKHETVVSYLEP-LSHIAA	278
HMM	SVWDMFWpLnGaTLyIpPeEtRrDPeKfWqYIqRHgITWwylTPSMFRM	
Query	279 QMMDIWVPIKIGALTYFAQADAL--KGTLVSTLKEVKPTVFIGVPQIWEK	326
HMM	LMpd.....	
Query	327 IHMVKKNSAKSMGLKKAFVWARNIGFKVNSKMLKYNTPVSYRMKT	376
HMM	.....psLRhVMFGepLsPehWdWWRkrfgkgRIINMYWPT	
Query	377 LVFSKVKTSGLGDHCHSFISGTAPLNQETAFFL-SLD--IPIGELYGLS	423
HMM	ETTVwtTwMrIiPdepeqWrwIPIGRPIpNtqWYIMDDnnMQLOPIGVIGE	
Query	424 ESSGPHTISNQNN--Y---RLLSCGKILTGCKNMLFQQN---KDG-IGE	463
HMM	LYIGWPGVARGYNNRPETTERFipNPFWPGEYRrGWNrRMYRTGDLAR	
Query	464 ICLWG-RHIFMGYLESETETETAIDEGW-----LHSGDLGQ	499
HMM	WLPDGNIEYLGRID.DQVKIRGYRIELGEIEhqlr.qHPgiqEAVV+	
Query	500 LDGLGFLYVTGHIKEILITAGGENVPPIPVETLVKKKIPIISNAML	545

DKFZphtes3\_35k24

group: transmembrane protein

DKFZphtes3\_35k24 encodes a novel 514 amino acid protein without similarity to known proteins.

The novel protein contains 5 transmembrane regions.

No informative BLAST results; No predictive prosite, pfam or SCOP motive.

The new protein can find application in studying the expression profile of testis-specific genes and as a new marker for testicular cells.

unknown ;

membrane regions: 5

Summary DKFZphtes3\_35k24 encodes a novel 514 amino acid protein. No homologues found in bacteria yeast and C.elegans, specific for mammals?

unknown

complete cDNA, complete cds, few EST hits

Sequenced by DKFZ

Locus: unknown

Insert length: 2706 bp

Poly A stretch at pos. 2696, polyadenylation signal at pos. 2675

```
1 CCGTGTGCAG TCGCCCCGCG CCCC CGCGGA CCCTTCGGGT AAACACGAA
51 CTGGGAGTTC TGAAGAATGG GTAAAGACTT TCGTTACTAT TTCCAGCATC
101 CCTGGTCTCG CATGATTGTG GCTTACTTGG TGATCTTCTT TAACCTCTTA
151 ATATTTCGGG AGGACCCAGT TTCTCATAGC CAAACAGAAG CCAATGTTAT
201 TGTGTGTGGA AACTGTTTTT CATTGTGTAC AAATAAATAC CCTAGAGGAG
251 TTGGCTGGAG GATTTTGAAG GTGCTTCTAT GGCTACTTGC CATTCTCACA
301 GGACTAATAG CTGGCAAATT TCTGTTCAT CAGCGTTTGT TTGTCAGTT
351 GCTCCGATTA AAAATGTTTC GAGAAGATCA TGGGTCTGGG ATGACAAATG
401 TCTTCAGCAC AATCTCTTTT CTCTTCATAT TTCTCAGAT ATACACACCG
451 ATTCTTCTAA TGGATGGGAA CATGGGAGCA TATATCATT CAGACTATAT
501 GCGCATCGGA AATGAAGATT TCATGAATTT AGCTGCAGTA GGGACCTGGA
551 TGGGGGACTT TGTACAGCTT TGGATGGTCA CTGATATGAT GCTTCAGGAC
601 AAACCCATAT CTGACTGGGG AAAATCAGCA AGAGCTTTCT GGAAGAAAGG
651 AAATGTTAGG ATCACTTTAT TCTGGACAGT TCTTTTACT CTGACGTCTG
701 TGGTTGTACT TGTGATTACA ACGGATCGGA TCAGCTGGGA CAAGCTGAAT
751 CGGGGATTTT TGCCCACTGA TGAAGTTTCC AGAGCATTCCT TTGCTTCTTT
801 TATCTTGGTC TTTGACCTTC TTAATTGTAT GCAGGACTGG GAATTCCTAC
851 ATTTTCATGG AGATGTTGAT GTAAATCTCC CTGGTTTGCA CACCCCTCAC
901 ATGCAGTTCA AGATCTCTTT CTTCAGAAA ATCTTCAAGG AGGAATATCG
951 TATTCACATA ACAGGCAAT GGTTAACATA TGGAAATATC TTCTCTGTCT
1001 TGATTTTGGA TCTTAATATG TCGAAGAAC AAATATTTTA TAAACCTCAT
1051 GAATATGGGC AATATATCGG CCGCGGGCAG AAGATATATA CAGTGAAAGA
1101 CTCAGAAAGT TTAAGAAGAT TGAACAGAAC CAAGCTATCC TGGGAATGGA
1151 GGTCCAATCA CACTAACCTT CGGACTAATA AAACATATGT TGAGGAGAGC
1201 ATGTTCTTAC ACAGCAGGTT CATAGGAGCC AGTCTTGATG TCAAGTGTCT
1251 GGCCTTTGTT CCAAGCCTGA TAGCCTTTGT GTGGTTTGGA TTCTTTATTT
1301 GGTTCCTTGG ACGATTTTGG AAAAATGAGC CACGCATGGA GAATCAAGAC
1351 AAAACTTACA CTCCGATGAA AAGAAAATCT CCATCAGAAC ATAGCAAGA
1401 CATGGGAATC ACTCGAGAAA ACACCCAGGC TTCAGTAGAA GACCCCTTGA
1451 ATGACCCCTC TTTGGTTTGC ATCAGGCTCG ACTTCAATGA GATCGTCTAC
1501 AAGTCTTCCC ACCTAACCTC GGAAACCTTG AGCTCAGAT TGAACGATC
1551 TACTAGTCCA ACAGAGCTG ATCAAGACCC AACGACTTCT AAACGTACAC
1601 CTACGACATA GACTCGGAGA TAGACTTGGA GATAACACAA AAGCAACCT
1651 TGAGTGAAC TTTAAAAATT TAGCTTTTCC TTTTGTATAT GTAAGTTTA
1701 CGTAGTGTTA GGTAAAAATA TGAACAAATG CACACGGGTG CTCACATGCG
1751 TTTTCTTAGG ATTCAATTGT TTCTATTGTG ATTATAATAC ACGTGCCTAC
1801 TGTAATCTCA ACAGTCTCTT AGAGATTGCT TTTCACAATT GCACAGCTAC
1851 TTACTGACTT TACAGCATAG TGGAGAGATTA GCTGATGACC CATGTATCTG
1901 ATGTTCAACC ATAGTGTGTC CTGAGACAT TAACTGTTT TTAAGTGTAC
1951 CAGAAATGAA GTGTGGAACA GTTACCTAAC CTATTTTACA TGGGCGTTTT
2001 GTATACAAC TTTTGTATCT ACACCTTGAT TCTGAGCAGA AAACAGAAAT
2051 AGCTAAATGT GACTCAGGAA GTATCTCTTG GTTTCTTATT CAGCAGCAGA
2101 GTTGGTGACT TTGACAACTG GACTGCAGAG AAACATGGTG ATCACCCTTT
2151 AATTTTTTAT GGCTGTCTGC CAATATATAA TACAGATGCA AATTCAGTA
2201 ATAGGAGATC CATACCCCAA CATGGGTAC TACTCGTGAA ATGTGACTTT
2251 CTCCACACAG TAATTGAAT GAGGTGATGA TACCTAATTA TGTTTCTCTA
2301 ATTAAGATA AATTGCTACT TGATTAAAA TCCTGCCCTT CACCTTTGGG
```

**PCT/IB00/01496**

2351	AACAAGGTT	AAGAGACACA	GTTGGGCGAA	CTCTCAATT	TATTGGCAAT
2401	TACAACAGT	CCGACAGCAT	CAGGACCAT	AGGTTTTCAT	CATTAGAGAG
2451	CTGCTGCTG	CAAGTTCGTA	ATATTCGTA	TATTTCTGT	CAAAATATGCC
2501	TCGTGGATG	GAANAATTGA	AAACATATGT	CAACCTGTG	CAGGGAAGAT
2551	CTCCAAAAA	TCATGACGGC	GAACTTCGT	AGGTGAGAGA	CGCGTGATG
2601	AAAGAATTG	TAAATCGT	TGTTTTCGT	ATGTTTGTT	TGTTTGTGT
2651	TTTATGAAG	TAAATATGTC	ACATTAAAT	ATAAGAATTA	TACAGCAAAA
2701	AAAAAA				

## BLAST Results

No BLAST result

### Medline entries

No Medline entry

Peptide information for frame 1

ORF from 67 bp to 1608 bp; peptide length: 514  
Category: putative protein

1	MGKDFRYKYP	PPMSRWIVAY	LWIFANFTGL	AEDCPVSHQST	EANVIVVGMN
51	FXGTRNYKYP	GRGMRILKVL	LWLAAILFTL	IACKPLFHQR	LFQALLQCLC
101	FREDHGSHMT	MFSFTSLFVL	IFSHYIITPL	LDMGNGMAYN	ITDYNGKNGN
151	SFKMLAAGVT	WGTDFEATMW	TDVMDLQDXT	YPWQKGSARA	YFWKGNVIRIT
201	LFVLVLTWMT	SVSVLVVITD	WISWOLHTRP	FLPSDEVSRA	FLASILVFLD
251	LTVLMQDFE	PHMGDOGVNV	LPGKLTPHMQ	KPIFFKFOPT	KEFYIRHIFIT
301	KWNRGTSIWL	VLTDLNLMNV	LOPFRKHEY	GOYIGPGOKI	YVYDSGSE
351	DLNVTGKLF	WRSHNTWVNV	WVWVWVWV	WVWVWVWV	WVWVWVWV
401	WVWVWVWV	WVWVWVWV	WVWVWVWV	WVWVWVWV	WVWVWVWV
451	WVWVWVWV	WVWVWVWV	WVWVWVWV	WVWVWVWV	WVWVWVWV
501	WVWVWVWV	WVWVWVWV	WVWVWVWV	WVWVWVWV	WVWVWVWV

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKF2phtes3\_35k24, frame 1

No Alert BLASTP hits found

Pedant information for DKFZphtes3\_35k24, frame 1

Report for DKFZphtes3\_35k24.1

[LENGTH]	514	
[MW]	60185.03	
[pI]	8.67	
[PROSITE]	MYRISTYL	5
[PROSITE]	CAMP_PHOSPHO_SITE	1
[PROSITE]	CK2_PHOSPHO_SITE	8
[PROSITE]	TYR_PHOSPHO_SITE	1
[PROSITE]	PKC_PHOSPHO_SITE	7
[PROSITE]	ASN_GLYCOSYLATION	6
[KW]	SIGNAL PEPTIDE 32	
[KW]	TRANSMEMBRANE 5	
[KW]	LOW COMPLEXITY	15.37

[illegible]

[illegible]

Prosites for DKFZphtes3 35k24.1

PS000001	149->153	ASN GLYCOSYLATION	PDOC000001
PS000001	353->357	ASN GLYCOSYLATION	PDOC000001
PS000001	363->368	ASN GLYCOSYLATION	PDOC000001
PS000001	371->375	ASN GLYCOSYLATION	PDOC000001
PS000001	487->491	ASN GLYCOSYLATION	PDOC000001
PS000001	493->497	ASN GLYCOSYLATION	PDOC000001
PS000004	435->439	CAMP PHOSPHO SITE	PDOC000004
PS000005	56->58	PKC PHOSPHO SITE	PDOC000005
PS000005	187->190	PKC PHOSPHO SITE	PDOC000005
PS000005	299->302	PKC PHOSPHO SITE	PDOC000005
PS000005	342->345	PKC PHOSPHO SITE	PDOC000005
PS000005	348->351	PKC PHOSPHO SITE	PDOC000005
PS000005	370->370	PKC PHOSPHO SITE	PDOC000005
PS000005	507->510	PKC PHOSPHO SITE	PDOC000005
PS000006	38->42	CK2 PHOSPHO SITE	PDOC000006
PS000006	342->346	CK2 PHOSPHO SITE	PDOC000006
PS000006	348->352	CK2 PHOSPHO SITE	PDOC000006
PS000006	373->377	CK2 PHOSPHO SITE	PDOC000006
PS000006	438->442	CK2 PHOSPHO SITE	PDOC000006
PS000006	456->460	CK2 PHOSPHO SITE	PDOC000006
PS000006	497->501	CK2 PHOSPHO SITE	PDOC000006
PS000006	499->503	CK2 PHOSPHO SITE	PDOC000006
PS000007	326->334	TYR PHOSPHO SITE	PDOC000007
PS000008	48->54	MYRISTYL	PDOC000008
PS000008	105->109	MYRISTYL	PDOC000008
PS000008	106->112	MYRISTYL	PDOC000008
PS000008	134->140	MYRISTYL	PDOC000008
PS000008	159->165	MYRISTYL	PDOC000008

(No Pfam data available for DKFZphtes3 35k24.1)



DKFZphtes3\_35n12

group: metabolism

DKFZphtes3\_35n12 encodes a novel 315 amino acid protein with strong similarity to ADP,ATP carrier T (ANT) proteins.

The novel protein contains three mitochondrial energy transfer signatures and is closely related to the ADP/ATP translocator, or adenine nucleotide translocator (ANT), a protein most abundant in mitochondria. In its functional state, it is a homodimer of 30-kD subunits embedded asymmetrically in the inner mitochondrial membrane. The dimer forms a gated pore through which ADP is moved from the matrix into the cytoplasm.

The new protein can find application in modulation of ADP-transport and energy metabolism in cells/mitochondria.

strong similarity to ADP/ATP carrier proteins

EST hits to mouse and drosophila

Sequenced by DKF2

Locus: unknown

Insert length: 1803 bp

Poly A stretch at pos. 1793, polyadenylation signal at pos. 1772

```
1 AGCGTCCCAA GAGCCACTTT CTCGCCAGTA CGATGCTGCA CCGGTTTTCC
51 GGTTCCTCCG TCCTCTTCAT CGTAGCTCCC GTACTCATTT TTAGCCACTG
101 CTGGCGGTTT TTATATCCCT CTCATCATG CATCGTGAGC CTGCGAAAAA
151 GAAGCGAGAA AAGCGGCTGT TTGACGCCCTC ATCCTTCGGG AAGGACCTTC
201 TGGCCGGCGG AGTCGCGGCA GCTGTGTCCA AGACAGCGGT GGCGCCCATC
251 GAGCGGGTGA AGCTGCTGCT GCAGGTGCGA GCGTCGTCGA AGCAGATCAG
301 CCCCAGGGCG CGGTACAAAG GCATGGTGGA CTGCTGGTG CGGATTCCCTC
351 GCAGCAGGGG TTTCTTCACT TTTTGGCGTG GCAATTTGGC AAATGTTATT
401 CGGTATTTTC CAACACAAGC TCTAAACTTT GCTTTTAAGG ACAAATACAA
451 GCAGCTATTC ATGCTGGGAG TTAATAAAGA AAAACAGTTC TGGAGGTGGT
501 TTTTGGCAAA CCTGGCTTCT GGTGGAGCTG CTGGGGCAAC ATCCTTATGT
551 GTAGTATATC CTCTAGATTT TGCCCGAACC CGATTAGGTG TCGATATTGG
601 AAAAGTCTCT GAGGAGCGAC AATTCAGGG TTAGGTGAC TGTATTATGA
651 AATAGCAAAA ATCAGATGCA ATTGCTGTT TATACCAAGG GTTTGGTGTT
701 TCACTACAGG GCATCATTTG GTACCGAGCC TCTTATTTG GAGCTTATGA
751 CACAGTTAAG GGTATTATAC CAAAGCCAAA GAAACTCCA TTTCTTGCTT
801 CCTTTTTCAT TGCTCAAGTT GTGACTACAT GCTCTGGAAT ACTTCTTAT
851 CCCTTTTGACA CAGTTAGAAG ACGTATGATG ATGCAGAGTG GTGAGGCTAA
901 ACGGCAATAT AAAGGAACCT TAGACTGCTT TGTGAAGATA TACCAACATG
951 AAGGAATCAG TTCTTTTTTT CGTGGCGCCT TCTCCAATGT TCTTCGCGGT
1001 ACAGGGGGTG CTTTGGTGTG GGTATTATAT GATAAATTA AAGAATCTT
1051 TCATATTGAT ATTGGTGSTA GGTAAATCGG AGAGTAAAT AAGAAATAAC
1101 ATGGATTTAA CTTGTTAAAC ATACAAATTA CATAGCTGCC ATTTGCATAC
1151 ATTTTGATAG TGTATTGTC TGTATTTGT TAAAGTGCTA GTTCTGCAAT
1201 AAAGCATACA TTTTTCACAG AATTTAAATA CTAAATCA GATAAATGTG
1251 GATTTTCCTC CCACTTAGAC TCAACACAT TTAGTGTA TATTTCATT
1301 ATTATAGGTA GTATATTTTA ATTTGTAGT TTAATTTCT TTTTATGATT
1351 AAAAATTAAT CATATAATCC TAGATTATG CTGAAATCTA GGAAATGAAA
1401 GTAGCGTCTT TTAATTTGCT ATTCATTAA TATACCTGTT TTCCCATCTT
1451 TTGAAGTCAT ATGGTATGAC ATATTCTTA AAAGCTTATC AATAGATGTC
1501 ATCATATGTG TAGGCAGAAA TAAGCTTTGT TCTATATCTC TTCTAAGACA
1551 GTGTATTATA CTGGTATATA TATTTACAGT ATCAGCCTTT GATTATAGAT
1601 GTGATCATTT AAAATTGAT AATGACTTTA GTGACATTAT AAAACTGAAA
1651 CTGGAAAAATA AAATGGCTTA TCTGCTGATG TTTATCTTTA AAAATAATAA
1701 AATCTTGCTA GTGTGAATAT ATCTTAGAAC AAAAGGTATC CTCTTGAAAA
1751 TTAGTTTGTA TATTTGTTG ACAATAAAGG AAGCTTAAC TTTAAAAAAA
1801 AAA
```

## BLAST Results

No BLAST result

## Medline entries

96289608:  
Molecular biological and quantitative abnormalities of  
ADP/ATP carrier protein in cardiomyopathic hamsters.

## Peptide information for frame 2

ORF from 128 bp to 1072 bp; peptide length: 315  
Category: strong similarity to known protein  
Classification: Metabolism  
Prosite motifs: MITOCH\_CARRIER (40-50)  
MITOCH\_CARRIER (145-155)  
MITOCH\_CARRIER (242-252)

1 MHREPARKKA EKRLFDASSF GKDLLAGGVA AAVSKTAVAP IERVKLLQV  
51 QASSKQISPE ARYKGMVDCL VRIPREQQFF SFWRGNLANV IRYFPTQALN  
101 FAFKDKYKQL FMSGVNKEKQ FWRWFLANLA SGGAGATSL CVVPLDFAR  
151 TRLCVDIGKG PEERQFKGLG DCINKIAKSD GIAGLYQGFV VSVQGIIVYR  
201 ASYFGAYDTV KGLLPKPKKT PFLVSFFIAQ VVTTCSGILS YPFDTVRRRM  
251 MQSGGEAKRQ YKGTLDCEVK IYQHEGSISS FRGAFSNVLR GTGGALVLLV  
301 YDKIKEFFHI DIGGR

## BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphtes3\_35n12, frame 2

PIR:S37210 ADP,ATP carrier protein T1 - mouse, N = 1, Score = 1127, P =  
2.7e-114

PIR:A44778 ADP,ATP carrier protein T1 - human, N = 1, Score = 1125, P =  
4.4e-114

TREMBL:DNADPATPT\_2 product: "ADP/ATP translocase"; Drosophila  
melanogaster gene encoding ADP/ATP translocase, N = 1, Score = 1124, P =  
5.6e-114

PIR:XWBO ADP,ATP carrier protein T1 - bovine, N = 1, Score = 1121, P =  
1.2e-113

>PIR:S37210 ADP,ATP carrier protein T1 - mouse  
Length = 298

## HSPs:

Score = 1127 (169.1 bits), Expect = 2.7e-114, P = 2.7e-114  
Identities = 214/293 (73%), Positives = 248/293 (84%)

Query: 17 ASSEKDLLAGGVA AAVSKTAVAPIERVKLLQVQASSKQISPEARYKGMVDCLVRIPRE 76  
A SF KD LAGG+AAAVSKTAVAPIERVKLLQVQ+SKQIS E +YKG++DC+VRIP+E  
Sbjct: 5 ALSFLKDLFAGGIAAAVSKTAVAPIERVKLLQVQHASKQISAEKQYKGIIDCVVRIPKE 64  
Query: 77 QGFFSFWRGNLANVIRYFPTQALNFAFKDKYKQF+GV++KQFWR+F NLASGGAAG 136  
QGF SFWRGNLANVIRYFPTQALNFAFKDKYKQF+GV++KQFWR+F NLASGGAAG  
Sbjct: 65 QGFLSFWRGNLANVIRYFPTQALNFAFKDKYKQIFLGGVDRHKQFWRIFAGNLASGGAAG 124  
Query: 137 ATSLCVVYPLDFARTRLCVDIKGPEERQFKGLGDCINKIAKSDGIAGLYQGFVSVQGI 196  
ATSLC VYPLDFARTRL D+CKG +R+F GLGDC+ KI KSDG+ GLYQGF VSVQGI  
Sbjct: 125 ATSLCFVYPLDFARTRLAADVKGSSQREFNGLGDCCLKIFKSDGLKGLYQGFVSVQGI 184  
Query: 197 IVYRASYPGAYDTVKGLLPKPKKT PFLVSFFIAQVVTTCSGILSYFPDTRRRMMMSQGE 256  
I+YRA+YFG YDT KG+LP PK +VS+ IAQ VT +G++SYFPDTRRRMMMSQGE  
Sbjct: 185 IIVRAAYFGYDTAKGMLPDPKNVHIIVSWMIAQSVTAVAGLVSYFPDTRRRMMMSQGR 244  
Query: 257 --AKRQYKGTLDCEVKIYQHEGSISSFRGAFSNVLRGTGGALVLLVYDKIKEF 307  
A Y GTLDC+ KI + EG ++FF+GA+SNVLRG GGA VLVLVD+IK++  
Sbjct: 245 KGADIMYGTLDCEWRKIADUEGANAFFRGAWSNVLRGAGAFVLLVYDEIKKY 297

## Pedant information for DKFZphtes3\_35n12, frame 2

Report for DKFZphtes3\_35n12.2

[LENGTH] 315

[illegible]

## Prosites for DKFZphtes3\_35n12.2

PS00215	40->50	MITOCH_CARRIER	PDOC00189
PS00215	145->155	MITOCH_CARRIER	PDOC00189
PS00215	242->252	MITOCH_CARRIER	PDOC00189

## Pfam for DKFZphtes3\_35n12.2

HMM_NAME	Mitochondrial carrier proteins		
HMM	*pFwkDFLAGGIAGMeHTvMFPIDtIKTRMOLQgEMpH...shpRYkGMI +F+KD+LAGG+A++++T+++PI+++K+++Q+O +++ RYKGM+		
Query	19	SFGKDLLAGGVAAVSKTAVAPIERVKLLQVQASSKQISPEARYKGMV	67
HMM	dCFRwIwkNEGWRLWRGLgANvIRYIPgWAIrFGFYEFMKEMfIDyfgE DC+ +I++++G++++WRG++ANVIRY+P++A++F+F++ +K +F + +++		
Query	68	DCLVRIPREQGFSSFWGNLANVIRYFPTQALNFAFKDKYKQLFMVGYNK	117
HMM	ddnyWmWfMnYMaGsmAGewIsvIitYPMWvVKTRLQaDqkHphsQp.R ++W+WF+ N+++G++AG ++S+ ++YP+++++TRL D +++++ R		
Query	118	EKQFWRWFLANLASGGAAG-ATSLCVVYPLDFARTRLGVD--IGKGPZER	164
HMM	hYNGvWncWrkIYReEGgFkGLYRGWtPTWNRMIpYqmiYFf+YEtLKeW +++G+ +C KI +++G ++GLY+G++ +++++I+Y++ YF++Y+T K +		
Query	165	QFKGLGDCIMKIAKSDG-IAGLYQGFVSVQGIIVYRASYPGAYDTVKGL	213
HMM	lynYtgYnPgprelCMddsPwWhwLIgWmIAGMiaWivSYpFDVVTRMM L +++ + ++++++I++ ++ +++++I+SYpFD+VR+RMM		
Query	214	LP-----KPK--KTPFLVSFFIAQVVT-TCSGILSYpFDVRRMM	251
HMM	MdsM.edhkYqSmLDChMqIYKneGFkGFwKGFwPRIMRiMPWtAIMPmI M+S+ +++++Y+++LDC+++IY++EG+ +F++G+ +++++R+ ++A++++		
Query	252	MQSGEAKROYKGTLDCEVKIYQHEGISFFRGAfSNVLRGT-GGALVLVL	300
HMM	YEqMKwFL* Y+ +K+F+		
Query	301	YDKIKEFF	308

DKFZphtes3\_35n24

group: testes derived

DKFZphtes3\_35n24 encodes a novel 365 amino acid protein without similarity to known proteins.

The novel protein contains a Prosite Ig(Immunoglobulin)-MHC pattern. This pattern represents a domain, approximately one hundred amino acids long and including a conserved intra-domain disulfide bond (Ilg domain1). Thus, the novel protein is a new member of the Ig-superfamily. No informative BLAST results; No predictive prosite, pfam or SCOP motif.

The new protein can find application in studying the expression profile of testis-specific genes.

unknown

complete cDNA, complete cds, EST hits

Sequenced by DKFZ

Locus: unknown

Insert length: 1589 bp

Poly A stretch at pos. 1579, polyadenylation signal at pos. 1560

```
1 CGATCGTCAC GTGACGCCGG GGTTCAGCGT ATCCTTGCTG GGCAACCGTC
51 TTAGAGACCA GCACTGCTGG CTGCACCATG AATGTGATCT ACCCACTGGC
101 AGTCCCAAG GGGCGCAGAC TCTGCTGTGA GGTGTGGGAA GCGCCAGCCG
151 AGCGGGTGTG CGCGGCCTGC ACAGTCACCTT ATTACTGTGG GGTGGTACAT
201 CAGAGGGCTG ACTGGACAG CATCCATGAG AAAATATGTC AGCTCTTGAT
251 TCCACTGGCG ACTTCCATGC CCTTCTACAA TTCAGAGGAA GAACGGCAGC
301 ATGGCCCTGA CGAGCTGCAG CAGCGGCAGA AGTATTTGAT TGAATCTGTC
351 TACACCATAG CCCAGAAATA CCTCTTTGAA GGGAAACAGC AAGATGCTGT
401 ACCAGCAGCT TTGCAGTCCC TTCGCTTCGG TGTGAAGCTG TATGGCCTGA
451 GCTCCGTAGA GCTTGTGCCT GCTTACCCGC TGTGGCCGA GCGCCAGCCTT
501 GGTCTGGGCC GAATCGTTCA GGCTGAAGAA TATCTATTCC AAGCCAGTGT
551 GACAGTCCCT AAATCAACTG ACTGTAGTAA TGCCACCCAC TCTTTACTGC
601 ATCGGAATCT GGGACTTCTC TATATAGCTA AGAAAACTA TGAAGAGGCC
651 CGTTATCATC TGCCCAATGA TATTTATTTT GCCAGTGTG CATTGGGAAC
701 AGAGGACATT AGGACTTCAG GAGGCTACTT CCACCTGGCT AATATATTCT
751 ATGACCTTAA AAAGTTGGAC CTGGCAGACA CATTGTACAC CAAGGTCTCT
801 GAGATCTGGC ATGCATATTT GAACAATCAC TATCAAGTCC TCTCAGGCG
851 TCACATCCAA CAAATGGATT TACTGGGCAA ACTATTTGAC AATGACACTG
901 GCTTGGATGA AGCCCAAGAA GCAGAAGCCA TTCGCATCCT GACTTCAATC
951 TTGAACATTC GAGAATCTAC ATCTGACAAA GCGCCCAAAA AAACCATCTT
1001 TGTTCTGAAG ATCCTGGTCA TGCTTTACTA CCTGATGATG ARTTCTTCAA
1051 AGGCACAGGA ATATGGCATG AGGGCCCTCA GTCTAGCCAA AGAACACNG
1101 CTTGATGTCC ATGAGCAAG CACCATTCAA GAGTATTAA GTCTCATTTC
1151 AACTGAGAC CATCCCATTA CTTAGTGACC CATGAGCTCT GCATCAAGGG
1201 TTATTCCAGG GGCTACTGAA GATCTAATAT ATTCCAGCCT TGCACAACGT
1251 CTTTGAGGTA CTGTAGACTG CTGAAGTTTC CACCCTCTTC CCCTGGGATT
1301 GCACACATAG CTGTTATTTT TTTCTTACAC AGCATATTAA GGAATATATA
1351 AGCTTTAGGC ATAGAAATCA CTAAAAACTG TGTTTGTCAAT GACCTTTGTA
1401 CTTGATTAT CATGACTTTG TATGACTGAG TAATATGTAG TCAGATCACT
1451 AATATGGTAT TTGTAATTAA ACTACAAATA GTTTGTGATT TCCCAAGAGT
1501 CTTCAACGA TGCTATGTTT ATACACTTTT GCTAAAGGAG GGGTAAAGGA
1551 GGGGGTAGGG AATAAAGCTA TATTGGAACA AAAAAA
```

## BLAST Results

No BLAST result

## Medline entries

No Medline entry

## Peptide information for frame 3

ORF from 78 bp to 1172 bp; peptide length: 365  
Category: putative protein

Prosite motifs: IG\_MHC (35-42)

```
1 MNVIYPLAVP KGRRLCCEVC EPAERVCAA CTVYYCGVV HOKADWDSIH
51 EKICQLLIPL RTSMPFYNSE EERQHLQOL QORQKYLIEF CYTIAQKYLIF
101 EGKHEDAVPA ALQSLAFRVK LYGLSSVELV PAYPLAEAS LGLGRIVQAE
151 EYLFQAQMTV LKSTDCSNAT HSLLRNLGL LYIAKKNYEE ARYHLANDIY
201 FASCAFGTED IRTSGGYFHL ANIFYDLKKL DLADTLTKV SEIWHAYLNN
251 HYQVLSQAH IQQMDLLGKLF ENDTGLDEAQ EAEAIRILTS ILNIRESTSD
301 KAPQKTIFVL KILVMLYYLM MNSSKAQEGY MRALSLAKEQ QLDVHEQSTI
351 QELLSLISTE DHPIT
```

## BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphtes3\_35n24, frame 3

No Alert BLASTP hits found

Pedant information for DKFZphtes3\_35n24, frame 3

## Report for DKFZphtes3\_35n24.3

```
[LENGTH] 365
[MW] 41768.24
[pI] 5.82
[BLOCKS] BL00273 Heat-stable enterotoxins proteins
[PROSITE] MYRISTYL 1
[PROSITE] IG_MHC 1
[PROSITE] AMIDATION 1
[PROSITE] CK2_PHOSPHO_SITE 7
[PROSITE] TYR_PHOSPHO_SITE 4
[PROSITE] PKC_PHOSPHO_SITE 3
[PROSITE] ASN_GLYCOSYLATION 3
(KW) Alpha_Beta
(KW) LOW_COMPLEXITY 4.11 %
```

```
SEQ MNVIYPLAVPGRRLCCEVC EPAERVCAA CTVYYCGVVHOKADWDSIHEKICQLLIPL
SEG .....
PRD ccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc

SEQ RTSMPFYNSEERQHLQOLQORQKYLIEFCYTIQKYLIFEGKHEDAVPAALQSLRFRVK
SEG .....
PRD ccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc

SEQ LYGLSSVELVPAYPLAEASLGLGRIVQAEYLFQAQMTVLKSTDCSNATHSLLRNLGL
SEG .....
PRD hhcccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc

SEQ LYIAKKNYEEARYHLANDIYFASCAFGTEDIRTSGGYFHLANIFYDLKKLADTLTKV
SEG .....
PRD eeehhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh

SEQ SEIWHAYLNNHYQVLSQAH IQQMDLLGKLFENDTGLDEAQEAEAIRILTSILNIRESTSD
SEG .....
PRD hhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhcc

SEQ KAPQKTIFVLKILVMLYYLMNNSKAQEGYMRALSLAKEQQLDVHEQSTIQELLSLISTE
SEG .....
PRD ccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc

SEQ DHPIT
SEG .....
PRD ccccc
```

## Prosite for DKFZphtes3\_35n24.3

PS00001	168->172	ASN_GLYCOSYLATION	PDOC00001
PS00001	272->276	ASN_GLYCOSYLATION	PDOC00001
PS00001	322->326	ASN_GLYCOSYLATION	PDOC00001
PS00005	114->117	PKC_PHOSPHO_SITE	PDOC00005
PS00005	299->302	PKC_PHOSPHO_SITE	PDOC00005
PS00005	323->326	PKC_PHOSPHO_SITE	PDOC00005

WO 01/12659

PCT/IB00/01496

PS00006	48->52	CK2_PHOSPHO_SITE	PDOC00006
PS00006	69->73	CK2_PHOSPHO_SITE	PDOC00006
PS00006	125->129	CK2_PHOSPHO_SITE	PDOC00006
PS00006	274->278	CK2_PHOSPHO_SITE	PDOC00006
PS00006	297->301	CK2_PHOSPHO_SITE	PDOC00006
PS00006	349->353	CK2_PHOSPHO_SITE	PDOC00006
PS00006	358->362	CK2_PHOSPHO_SITE	PDOC00006
PS00007	85->93	TYR_PHOSPHO_SITE	PDOC00007
PS00007	186->194	TYR_PHOSPHO_SITE	PDOC00007
PS00007	186->194	TYR_PHOSPHO_SITE	PDOC00007
PS00007	185->194	TYR_PHOSPHO_SITE	PDOC00007
PS00008	275->281	MYRISTYL	PDOC00008
PS00009	11->15	AMIDATION	PDOC00009
PS00290	35->42	IG_MHC	PDOC00262

(No Pfam data available for DKF2phtes3\_35n24.3)

DKFZphtes3\_35n9

group: metabolism

DKFZphtes3\_35n9 encodes a novel 607 amino acid protein which is a splice variant of human carboxylesterase (EC 3.1.1.1).

The novel protein contains both, one carboxylesterase B1 and one B2 pattern. In comparison to EC 3.1.1.1, DKFZphtes3\_35n9 shows a N-terminal extension and aa 458-474 are missing.

The new protein can find application in modulation of carboxylester metabolism and as a new enzyme for biotechnologic production processes.

carboxylesterase, splice variant

5' extension of mRNA and N-terminal elongation of protein (64 aa), missing exon! aa 458-474 of JC5408 are missing

Sequenced by DKFZ

Locus: unknown

Insert length: 2888 bp

Poly A stretch at pos. 2878, no polyadenylation signal found

```
1 CTCGGCCTGA GGTGCGAGAG AAGCGGTGAC CGCGGCCCTG GCTGCTCGGA
51 CCCGGGAACA TGATGGTTCG TGGAGCAGAA GGCCTGAGA AGGGACCACG
101 GCGGCGCTGG GTCGTGCGAG CCAGTAGCGG GCTGAAACGT AGAGGCCAGA
151 ACCAGGTCTC AGGGGGCACT AAAGGCGGTC GGAGGTAATC CCCACACCGC
201 TTCCTCTCTG AAGTCAGGCT GCGCGGGAGC TCCCGTATCC AGGACGTTG
251 GTCGGCTCTG GCTCGGAGG GATCTGAGTG TCTCGGGAGC TCCTGCTGAC
301 GCGCCTGCCT CCCCTGCTGC ACATAGGCG CGGAGTAGCG GCGTCCCCAC
351 AGCTTGGACC GGCAGGGGCT CGTGAATGT TTGTCAATG GATAAATGAC
401 CATGGCCGTG GTCTCCGGCG GAGGTGAGGA AACTGAAAGC CACCGAGGAA
451 AAGGGGGCGG CTCCTTAAGA AGTGCCCGCG TCACGTGTAC GTTTCAAAAG
501 AATGGCGTGA CTGAGTAGGG AGGGGACCGG GGAGACCTTC AGACCCCTGA
551 CTGTAAGGAG ATGAGGGGCC GTGAAGGGGA ACCCAGGAAA CTGAGTCTCTG
601 AAAGCAAGGA GGAACCTCCA GAATGAAGG CGCGGAGACT CCTTCTGCGC
651 TTGCTCAAG CGGTCTCTC ACCCGATCA AGTTCCTTCC CATTTCTCCA
701 TCTGGGGAT CCGTAACGTG CACATCTCA GAGAAGCCCT CCGGGGTCT
751 CCAATTCTAG TTTATTGCC CCTCTATCG ATCCCCAGC GCGCTCATCG
801 GGGCTGTGGA CAAGGACAGG TTTGAAGAGA GGATTCCTGT GATCGCGGAA
851 GGGCTGCAGG AATGGCACAG CCCCTTCCGA GGATGCCAAA GGAGCCCGGG
901 CAAAGGAAAG TGGCCGTGCC CGGGCCTGCC TACCCTAGA TCCCAACCCA
951 CCTATGACTG CTCAGTCCCG CTCTCTTACC ACACCCACCT TTCCGGGCCC
1001 AAGCCAGGCG ACCCGCTGA CTCCCTGCC AGTCCAACT CCAAGGCTGG
1051 GCAAGGCACT GATCCACTGC TGGACAGACC CGGGGAGCCG TCTGGGTGAA
1101 CAGCAGCGTG TCCGCCGGCA GCGAACCGAG ACCAGCGAGC CGACCATGCG
1151 GCTGCACAGA CTTCGTGCGC GGCTGAGCGG GGTGGCCTGT GGGCTTCTGC
1201 TGCTTCTTGT CCGGGGCCAG GGCCAGGACT CAGCCAGTCC CATCCGAGCC
1251 ACACACACGG GGCAGGTGCT GGGGAGTCTT GTCCATGTGA AGGCGGCCAA
1301 TGCCGGGGTC CAAACCTTCC TGGGAATTCC ATTTCGCAAG CCACCTCTAG
1351 GTCCGCTGCG ATTTCGACCC CCGTGAACCC CTGAATCTTG GAGTGTGTCT
1401 AGGAGTGAA CCAACCATCC GGCCTGTGT CTACAGGACC TCACCGCAGT
1451 GGACTCAGAG TTTCTTAGCC AGTTCAACAT GACCTTCCCT TCGACTCCA
1501 TGCTGAGGA CTGCTGTATC CTCAGCATCT ACACCGCGCG CCATAGCCAT
1551 GAAGGCTCTA ACCTGCCGCT GATGTTGTTG ATCCACGGTG GTGCGCTTGT
1601 TTTTGGCATG GCTTCCTTGT ATGATGGTTC CATGCTGGCT GCCTTGGAAG
1651 ACGTGGTGGT GGTCAATATC CAGTACCGCC TGGGTGTCTT GGGCTTCTTC
1701 AGCACTGGAG ACAAGCACGC AACCGGCAAC TGGGGCTACC TGGACCAAGT
1751 GGCTGCACCTA CGCTGGGTCC AGCAGAATAT CGCCCATTTT GGAGGCAACC
1801 CTGACCTGTG CACCATTTTT GCGGATCTG CGGGTGGCAC GAGTGTGTCT
1851 TCGCTTGTG TGTCCCCCAT ATCCCAAGGA CTCTTCCAGC GAGCATCAT
1901 GGAGAGTGGC GTGGCCCTCC TGCCCGGCTT CATTGCCAGC TCAGCTGATG
1951 TCATCTCCAC GGTGGTGGCC AACCTGTCTG CCTGTGACCA AGTTGACTCT
2001 GAGGCCCTGG TGGGCTGCCT GCGGGGCAAG AGTAAAGAGG AGATTCTTGC
2051 AATTAAACAAG CCTTCAAGA TGATCCCCGG AGTGTGGAT GGGGTCTTCC
2101 TGCCCAAGGA CCCCCAGGAG CTGCTGGCTT CTGCGACTT TCAGCCTGTG
2151 CCTAGCATTG TTGGTGTCAA CAACAATGAA TTCGGCTGGC TCATCCCAA
2201 GGTCAAGAG ATCATGATA CCGAGAAGA AATGGAGACA GAGGCTCTCC
2251 AGGCTGCTCT GAGAAAATG TTAACGCTGC TGATGTTGCC TCCTACATTT
2301 GGTGACCTGC TGAGGAGGGA GTACATTGGG GACAATGGGG ATCCCCAGAC
2351 CCTCAAGCGC CAGTTCAGG AGATGATGGC GGAATCCATG TTTGTGATCC
2401 CTGCACTCCA AGTAGCACAT TTTCAGTGT CCCGGGCCCC TGTACTTTC
2451 TACGAGTTCC AGCATCAGCC CAGCTGGCTC AAGAATATCA GGCACCGCA
2501 CATGAAGGCA GACCATGTTA AATTCATCTA GGAAGAGGAG CAGCTAAGCA
2551 GGAAGATGAT GAAGTACTGG GCAACTTTTG CGAGAAATGG GAGCCCAT
2601 GCGAGGGTGC TGCCACACTG GCGCTGTTC GACAGGAGG AGCAATACCT
```



WO 01/12659

PCT/IB00/01496

```
2651 GCAGCTGAAC CTACAGCCTG CGGTGGGCGG GGCTCTGAAG GCCCAGGCG
2701 TCCAGTTCTG GAAGAAGGCG CTGCCCCAAA AGATCCAGGA GCTCGAGGAG
2751 CTTGAAGAGA GACACACAGA GCTGTAGCTC CTTGTCCCGG GGAGGAGGGG
2801 CTGGGTTCCG TGACAGGCGA GGCTAGCCT GCTGTCCCA CACACACCA
2851 CTAAGGAGAA AGAAGTTCAT TCCTTCATAA AAAAAAAA
```

#### BLAST Results

Entry D50579 from database EMBL:  
Homo sapiens mRNA for carboxylesterase, complete cds.  
Score = 7197, P = 0.0e+00, identities = 1441/1443

Entry JC5408 from database PIR:  
carboxylesterase (EC 3.1.1.1) - human  
Score = 2808, P = 1.2e-291, identities = 542/559, positives = 543/559,  
frame +3

#### Medline entries

No Medline entry

#### Peptide information for frame 3

ORF from 954 bp to 2774 bp; peptide length: 607  
Category: known protein  
Classification: Metabolism  
Prosite motifs: CARBOXYLESTERASE\_B\_1 (279-295)  
CARBOXYLESTERASE\_B\_2 (185-196)

```
1 MTAQSRSPIT PTFPGPSQRT PLTPCPVOTP RLKALIHCV TDPGQPLGEG
51 QRVRNRQRTET SEPTNRLHRL RARLSAVACG LLLLVRGQG QDSASPIRTT
101 HTGQVLGSLV HVKANAGVQ TFLGIPFAKP FLGPLRFAPP EPPEWSGVR
151 DGTTHPAMCL QDLTAVESEF LSQFNMTFPS DSMSEDCLYL SIYTPAHSHE
201 GSNLPVMVMW HGGALVFGMA SLYDGSMLAA LENVVVVLIQ YRLGVLGFFS
251 TGDKHATGNW GYLDQVAALR WQQNIAHFG GNPDRVITFG ESAGGTSVSS
301 LUVSPISQGL FHGAIMESGV ALLPGLIASS ADVISTVVAN LSACDQVDSE
351 ALVGCLRGKS KEEILAINKP FRMIPGVVDG VFLPRHPQEL LASADFPQVP
401 SIVGVNNEF GWLIPKVMRI YDTQKEMDRE ASQAALQKML TLLMLPPTFG
451 DLLREYIGD NGDPTQLQAO FQEMMADSMF VIFALQVAF QCSRAPVIFY
501 EFOHOPSMWK NRRPPMKAD HVKFTEEEO LSRKMMKYWA NFARNGNPNG
551 EGLPHMPLFD QEEQYLQNL QPAVGRALKA HRLQFWKKAL PQKIQELEEP
601 EERHTEL
```

#### BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphtes3\_35n9, frame 3

PIR:JC5408 carboxylesterase (EC 3.1.1.1) - human, N = 1, Score = 2808,  
P = 1.9e-292

TREMBL:HSU60553\_1 gene: "hCE-2"; product: "carboxylesterase"; Human  
carboxylesterase (hCE-2) mRNA, complete cds., N = 1, Score = 2761, P =  
1.8e-287

PIR:A34329 60K esterase (EC 3.1.1.-) isoform 2 - rabbit, N = 1, Score =  
1985, P = 3.1e-205

TREMBL:D50580\_1 product: "carboxylesterase precursor"; Rattus  
norvegicus mRNA for carboxylesterase, partial cds., N = 1, Score =  
1984, P = 4e-205

>PIR:JC5408 carboxylesterase (EC 3.1.1.1) - human  
Length = 559

HSPs:

Score = 2808 (421.3 bits), Expect = 1.9e-292, P = 1.9e-292

Identities = 542/559 (96%), Positives = 543/559 (97%)

Query: 65 MRLHRLRLRLSAVACGLLLLVRCQQDSASPIRTTHTGVLGSLVHVKGANAGVQTFILG 124  
 Sbjct: 1 MRLHRLRLRLSAVACGLLLLVRCQQDSASPIRTTHTGVLGSLVHVKGANAGVQTFILG 60

Query: 125 IPFAKPPLGLPLRFAPPEPPESWSGVRDGTTHPAMCLQDLTAVESEFLSQFNMTPFSDSMS 184  
 Sbjct: 61 IPFAKPPLGLPLRFAPPEPPESWSGVRDGTTHPAMCLQDLTAVESEFLSQFNMTPFSDSMS 120

Query: 185 EDCLYLSIYTPAHSHEGSLNPMVMVHGGALVFGMASLYDGSMLAALENVVVVIQYRLG 244  
 Sbjct: 121 EDCLYLSIYTPAHSHEGSLNPMVMVHGGALVFGMASLYDGSMLAALENVVVVIQYRLG 180

Query: 245 VLGFSTGDKHATGNMGYLDQVAALRWVQONIAHFGGNPDRVTIFGESAGGTSVSSLVVS 304  
 Sbjct: 181 VLGFSTGDKHATGNMGYLDQVAALRWVQONIAHFGGNPDRVTIFGESAGGTSVSSLVVS 240

Query: 305 PISQGLFHGAIMESGVALLPGLIASSADVISTVVANLSACDQVDSALVGCLRGKSKKEE 364  
 Sbjct: 241 PISQGLFHGAIMESGVALLPGLIASSADVISTVVANLSACDQVDSALVGCLRGKSKKEE 300

Query: 365 LAINKPFKMPGVVDGVFLPHRPOELLASADFPVPSIVGVNNNEFGWLI PKVMRIYDTQ 424  
 Sbjct: 301 LAINKPFKMPGVVDGVFLPHRPOELLASADFPVPSIVGVNNNEFGWLI PKVMRIYDTQ 360

Query: 425 KEMDREASQAALQKMLTLMLPPTFGDLLREYIGDNGDPQTLQAQFOEMMADSMFVIPA 484  
 Sbjct: 361 KEMDREASQAALQKMLTLMLPPTFGDLLREYIGDNGDPQTLQAQFOEMMADSMFVIPA 420

Query: 485 LQVAHFQCSRAPVYFYEFQHQPSWLNIRPPHMKADH-----VKFTEE 528  
 Sbjct: 421 LQVAHFQCSRAPVYFYEFQHQPSWLNIRPPHMKADH +KFTEE 480

Query: 529 EQLSRKMKMYWAFARNGNPNHGEGLPHWPLFDQEEQYLQNLQPAVGRALKAHRLQFWKK 588  
 Sbjct: 481 EQLSRKMKMYWAFARNGNPNHGEGLPHWPLFDQEEQYLQNLQPAVGRALKAHRLQFWKK 540

Query: 589 ALPQKIQELEEPEERHTEL 607  
 Sbjct: 541 ALPQKIQELEEPEERHTEL 559

Pedant information for DKF2phtes3\_35n9, frame 3

Report for DKF2phtes3\_35n9.3

[LENGTH] 607  
 [MW] 67051.20  
 [PI] 6.11  
 [HOMOL] PIR:JC5408 carboxylesterase (EC 3.1.1.1) - human 0.0  
 [BLOCKS] BL01173A Lipolytic enzymes "G-D-X-G" family, histidine  
 [BLOCKS] BL00122G  
 [BLOCKS] BL00122F  
 [BLOCKS] BL00122E  
 [BLOCKS] BL00122D Carboxylesterases type-B serine proteins  
 [BLOCKS] BL00122C Carboxylesterases type-B serine proteins  
 [BLOCKS] BL00122B Carboxylesterases type-B serine proteins  
 [BLOCKS] BL00122A Carboxylesterases type-B serine proteins  
 [SCOP] dlakn\_ 3.56.1.1.4 Bile-salt activated lipase [Bovine (Bos taurus 1e-158  
 [SCOP] d2ack\_ 3.56.1.1.1 Acetylcholinesterase [Electric ray (Torped 1e-170  
 [SCOP] dlthg\_ 3.56.1.9.7 type-B carboxylesterase/lipase [fungu 1e-149  
 [EC] 3.1.1.13 Sterol esterase 1e-52  
 [EC] 3.1.1.7 Acetylcholinesterase 5e-74  
 [EC] 3.1.1.1 Carboxylesterase 0.0  
 [EC] 3.1.1.8 Cholinesterase 5e-68  
 [EC] 3.1.1.59 Juvenile-hormone esterase 1e-34  
 [EC] 3.1.1.3 Triacylglycerol lipase 3e-52  
 [PIRKW] duplication 2e-47  
 [PIRKW] homotetramer 3e-67  
 [PIRKW] transmembrane protein 9e-44  
 [PIRKW] microsome 1e-130  
 [PIRKW] pancreas 3e-52  
 [PIRKW] endoplasmic reticulum 1e-134  
 [PIRKW] homotrimer 1e-134  
 [PIRKW] phosphatidylinositol linkage 5e-74  
 [PIRKW] synapse 3e-73  
 [PIRKW] liver 1e-131  
 [PIRKW] heparin binding 3e-52

```

[PIRKW]      phosphoprotein 7e-25
[PIRKW]      glycoprotein 1e-134
[PIRKW]      thyroid hormone biosynthesis 2e-47
[PIRKW]      carboxylic ester hydrolase 0.0
[PIRKW]      monomer 2e-42
[PIRKW]      disulfide bond 2e-31
[PIRKW]      mammary gland 3e-52
[PIRKW]      alternative splicing 5e-74
[PIRKW]      iodine 2e-47
[PIRKW]      pyroglutamic acid 6e-39
[PIRKW]      hydrolase 1e-135
[PIRKW]      muscle 3e-73
[PIRKW]      thyroid gland 2e-47
[PIRKW]      membrane protein 3e-73
[PIRKW]      neurotransmitter degradation 3e-73
[PIRKW]      cholesterol 3e-52
[PIRKW]      homodimer 2e-47
[PIRKW]      nerve 3e-73
[SUPFAM]     cholinesterase 0.0
[SUPFAM]     triacylglycerol lipase 1e-32
[SUPFAM]     cholinesterase homology 0.0
[SUPFAM]     thyroglobulin 2e-47
[SUPFAM]     thyroglobulin type I repeat homology 2e-47
[SUPFAM]     juvenile-hormone esterase 2e-35
[SUPFAM]     probable lipolytic protein ybaC 1e-07
[PROSITE]    CARBOXYLESTERASE_B_2 1
[PROSITE]    CARBOXYLESTERASE_B_1 1
[PFAM]       Carboxylesterases
[KW]         Alpha_Beta
[KW]         3D
[KW]         LOW_COMPLEXITY      3.95 %

SEQ      MTAQSRSPPTTPTFGGPSQRTPLTPCPVOTPRLGKALHCWTDPGQPLGEQQRVRRQRTET
SEG      .....XXXXXXXXX...
1acj)- .....

SEQ      SEPTMRLHRLRLRLSAVACGLLLLVRGQGQDSASPIRTHTGQVLGSLVHVKGANAGVO
SEG      .....XXXXX...
1acj)- .....ETEEEECEEEEEETTEE--EE

SEQ      TFLGIPFAKPLPLGLRFAPPEPPESWSGVRDGTTHPAMCLQDLTAVESEFLSQFNMTFPS
SEG      EEEEECEETTGGGTTCCCECCCCCECCCCCCCCCCCCCTTTT-HHHHCCCCC
1acj)- .....

SEQ      DMSSEDCLYLSIYTPAHSHEGSLPVMVWIRGGALVFGMASLYDGSMLAALENVVVVIIQ
SEG      CCBTTTTCEEEEEET--TTTTTEEEEEECTTTTTCTTTTGCCHHHHHHCCCEEECC
1acj)- .....

SEQ      YRLGVLGFFSTGDKHATGNWGYLDQVAALRWVQNIHAFGGNPDRTVIFGESAGGTSVSS
SEG      CCCCCGGCCCTTTTTTCCCHHHHHHHHHHHHHHCGGGGCEEEEEEECHHHHHHHH
1acj)- .....

SEQ      LVVSPISQGLFHGAIMESGVALLPLGIASSADVISTVVANLSACQVDSALVGCLRGKS
SEG      HHHCGGGTTTTCEEEEEETTTTTTTTBCCHHHHHHHHHHHHCC-CCCCCHHHHHHHHHHCC
1acj)- .....

SEQ      KEEILAINKPFKMI PGVVGVLPRHPQELLASADFQVPVSI VGNNEFGWLI PKVMRI
SEG      HHHHHHHHTCCCTTTCBTTTTTTTTTHHHHHHTTCCCEEEEEETBTNNHHHTTTTT
1acj)- .....

SEQ      YDTQKENDREASQAALQKMLTLLMLPPTFGDLLREYIGDNGDPQLQAQFQEMMADSMF
SEG      TTTCCCCCHHHHHHHHHHTTTTCHHHHHHHHHCTTTTTTHHHH-NHHHHHHHHHHH
1acj)- .....

SEQ      VIPALQVAFQCSRAPVYFYEFOHQPSWLKNI RPPHMKADHVRFTEEEQLSRKMKYWA
SEG      HHHHHHHHHHHHCCCCCEEEECCECCGGGTBTTHHCGGGCCCHHHHHHHHHHHHHH
1acj)- .....

SEQ      NFARNGNPNGEGLPHWPLFDQEEQYLQNLQPAVGRALKARLQFWKKALPQKIQEELP
SEG      HHHHHCCCCC--CCCCBTBTBEEEECCCCCEETTTTHHHHHHHHHHHH.....
1acj)- .....

SEQ      EERHTEL
SEG      XXXXXX
1acj)- .....

```

Prosites for DKF2phtes3\_35n9.3

PS00122 279->295 CARBOXYLESTERASE\_B\_1 PDOC00112  
PS00941 185->196 CARBOXYLESTERASE\_B\_2 PDOC00112

Pfam for DKFZphtes3\_35n9.3

HMM_NAME	Carboxylesterases
HMM	*MfMnwllmFLlwaItWii.WhegaprPdPyivdtncGkIRGmNedtD + *L+++ ++++++ ++Q+++P I T+ G+G++ +
Query	69 RLRARLSAVACGLLLLVRGQGDSASP---IRTTHT-GQVLGSLVHK 113
HMM	NG..pYYvFLGIPYAEPVGNLRFKePQPyhePwtNVNATnYPpMCMQW + + +FLGIP+A+PP+G LRF +P+P +E W++V++ T+ P MC+Q+
Query	114 GANAGVQTFLGIPFAKPLGLRFAPPEP-PESWSGVRDGTTHPAMCLQD 162
HMM	ndFGFWlEdnierWHenip..eMSDECLYNVMTPhnrkPNskLPVMVMI +++ ++N++ P +MSDECLYL+++TP+ + +S+LPVMVMI
Query	163 LTAV--ESEFLSQFNMTFPSDSMSDECLYLSIYTPAHSHGSLNLPVMVMI 210
HMM	HGGGFmFGSGhsYFlIqYDgeylMeeNVIVvtINyRLGPFGLSTgDid HGG+++FG + ++YDG+ L++ ENV+VY I+YRLG++GF+STGD +
Query	211 HGGALVFGMA----SLYDGSMLAALENVVVIQYRLGVLFSTGDKH 255
HMM	lPPHGNWGLWDQRMALQWQDNIAFGGDPNNITIFGESAGGNSVHlHML + GNWG++DQ++AL+WVQ+NIA+FGG+P+++TIFGESAGG+SV+ ++
Query	256 AT--GNWGYLDQVAALRWVQONIAHFGGNPDRVTIFGESAGGTSVSSLV 303
HMM	SYGGDNPPmKqLFHRAIMQSGsAmcPwVIQsnyNaRqRArFrimGCN S P + +LFH AIN+SG A+ P++I S++ + +A+++ C+
Query	304 S-----PISQGLFHGAIMESGVALLPGLIASSA--DVISTVVANLSACD 345
HMM	rmDssEMiqCLRskPwEELmdAtWnFwmMfYFPFLPMFFgPVIDGDdape + DS++++ CLR K+ EE+++++ +F + + +DG+
Query	346 QVDSEALVGLRGKSKKEILAINK---PFKMPGV-----VDGV---- 381
HMM	aFIPDHPeemIKEGkFndVPWIIQYNnDEGIWFapMmnnfnWfdeDeWid F+P+HP+E++++ F VP I+G+NN E++W++P M + + +E++
Query	382 -FLPRHPQELLASADFQPVFSIVGVNNEFGWLIPKVMRIYDT-QKEMDR 429
HMM	itNedWyeWMPYILFYrddasNikDMDDYIdkvyEeYPgWMDrFPqESYW ++ + ++ M +L + + + D ++EEY+G+ + PQ
Query	430 EASQAALQKMLTLMLPPT-F-----GDLLREEYIGDNGD-PQTLQA 469
HMM	nLQDMFTDYLFWCPTrihadnHRKhgspVYMYeFDHppSFGYgQFFmWR ++Q+M+ D F++P + ++H++ +PVY+YEF+H PS +
Query	470 QFQENMADSNFVIP--ALQVAFQCSRAPVYFYEQHQPSW-----LKN 511
HMM	WWPpWmgvdH* +PP+M++DH
Query	512 IRPPHMKADH 521
HMM	*TEEEIissMRmHQYWINFAKhGNPNthnglCWpQYTenEQYdMIME TEEE+ +S R HM+YM+NFA++GNFN++ GL++RP +++EQY++ +
Query	525 TEEEEQLS-RKMHKYANFARNGNPNGE--GLPHWPLFDQEEQYLQLNL 570
HMM	tIImiOmCrmrDPYCNFW* + +++++ + FW
Query	571 QPAVGRALKARR--LQFW 586

DKFZphtes3\_35p17

group: testes derived

DKFZphtes3\_35p17 encodes a novel 505 amino acid protein with weak similarity to  
Proteins of the armadillo family.

Proteins of the armadillo family are involved in diverse cellular processes in higher  
eukaryotes. Some of them, like armadillo, beta-catenin and plakoglobins have dual functions in  
intercellular junctions and signalling cascades. Others, belonging to the importin-alpha-  
subfamily are involved in NLS recognition and nuclear transport, while some members of the  
armadillo family have as yet unknown functions. The novel protein shows similarity to S.  
cerevisiae protein Yel013p (VAC8) and Danio rerio b-catenin, but contains no armadillo (arm)  
repeats.

No informative BLAST results; No predictive prosite, pfam or SCOP motif.

The new protein can find application in studying the expression profile of testis-specific  
genes.

similarity to S.cerevisiae VAC8

complete cDNA, complete cds, few EST hits

Sequenced by DKFZ

Locus: unknown

Insert length: 1966 bp

Poly A stretch at pos. 1956, polyadenylation signal at pos. 1935

```
1 AAGTCAATG TAAGATTGGT TCATTAAAAA TACTGAAGGA AATCAGTCAT
51 AATCCTCAAA TCAGACAGAA TATTGTTGAC CTGCGGGGCT TACCAATTAT
101 GGTGAATATA CTTGATTCTC CACACAAGAG TCTAAATGT TTGGCAGCCG
151 AGACTATCGC GAATGTTGCC AAGTTTAAAA GAGCACGGCG GTTGGTGAGG
201 CAGCACGGGG GTATCACCAG ACTGGTTGCT CTACTAGACT GTGCACATGA
251 TTCCACAAAA CCTGCCCAAT CGAGTCTGTA TGAGGCCAGA GACGTGGAAG
301 TGCGTCGCTG TGGGGCACTG GCCCTGTGCA GCTGCATAA GACTCATAG
351 AATAAGAGAG CCATCCGCAA AGCTGGGGCG ATTCTCTGT TGGCTGGGCT
401 GCTGAAGACT TCTCATGAAA ACATGCTAAT TCCAGTGGTG GGGCATTGCG
451 AAGAGTGTGC ATCAGAGGAA AACTACCGGG CTGCAATCAA AGCAGAAAGG
501 ATCATTGAAA ACCTTGTCAG GAACCTAAAT AGTGAGAAAG AGCAGCTGCA
551 GGAGCACTGC GCCATGGCCA TTTACCAGTG TGCTGAAGAT AAGGAAACCC
601 GGGACCTCGT TAGGCTGCAC GGAGGACTTA AGCCCTTGGC CAGTCTACTC
651 AATAACACTG ACAATAAAGA GCGGTTAGCT GCTGTACAG GGGCTATATG
701 GAATGTTTCC ATCAGCAAG AGATGTTTAC CAAGTTTCGG GANTACAAG
751 CCATTGAAAC CTTGCTGGGA CTCTAACAG ATCAGCCTGA AGAAGTACTT
801 GTGATGTGGG TTGGGGCCTT GGGAGAATGC TGCCAAGAAC GTGAAAACCG
851 AGTCATTGTC CGGAATGTG GTGGCATTCA ACCACTTGTG AACCTCCTTG
901 TTGGAATAAA CCAAGCTCTT CTTGTGAATG TTACAAAAGC AGTTGGTGCT
951 TGTGAGTAG AACCTGAAAG TATGATGATA ATTGATCGCT TAGATGGAGT
1001 TCGTTTGTGG TGCTCCCTGC TGAAAAATCC TCACCCAGAC GTGAAGGCCA
1051 GCGCAGCATG GGCACCTCTG CCAATGATCA AAAATGCAA GATGCTGGG
1101 GAATGCTTCT GTTCTCTTGT TGGTGGTTTG GAACCTATTG TCAATTACTT
1151 GAAATCAGAT AACAAGAGAG TTCTGGCAAG TGTATGTGCT GCCATTACCA
1201 ACATAGCAAA AGATCAAGAA AATTGAGCTG TTATCAGAGA TCATGGAGTT
1251 GTTCTTTTAT TGTCCAAACT GGCAAAATCA AATAACAATA AATTGAGACA
1301 TCATCTAGCA GAAGCTATTT CACGTTGCTG TATGTGGGGC AGGAATAGAG
1351 TGGCCTTCGG TGAGCACAAA GCAGTGGGTC CACTAGTGCG TTATCTGAAA
1401 TCAATGACA CCAAGCTGCA TCGGGCGACA GCTCAGGCTT TGTACCACT
1451 CTCAGAAGAC GCCGATACCT GCATGACAT GCATGAGAT GGTGCAATA
1501 AGCTCTTACT GGATATGCTT GGGTCCCTTG ACCAGGATCT CCAGGAAGCT
1551 GCAGCTGGTT GTATATCCAA TATCCGCGAG CTGGCTCTTG CTACAGAGAA
1601 GGCAAGATAC ACTTGAAATT TAAATGGACA TTACAAGCTA TCAAAATTCTA
1651 CATGACACAG GACATGTCTC TCCCATGGCC AGAAAGCCTA AATTGGGAAA
1701 CAGTTGTTAG CAAACCCCTT CAACUATCTA AATGAAACA CACAAATTGA
1751 AATGACACAG AATGTTTTTC ATCTGAAAT TGCATGGAGA CTTTTGTTTC
1801 TATTTAATGT TTTGAGATA TGACATGTGA TAAGATGGAA AGCCAATAAA
1851 CCTGTGATTA GTTCTTAAGA ATATGAGAT ATAGTATAT GATGTATTTT
1901 TAGTCAAGTG ATGCTTTTGT ATTTGTGGCG ATTTTAATAA AGGATATGGC
1951 CTTCCCAAAA AAAAAA
```

## BLAST Results

No BLAST result

## Medline entries

98413148:  
Yel013p (Vac8p), an armadillo repeat protein related to plakoglobin and importin alpha is associated with the yeast vacuole membrane.

98330438:  
YEB3/VAC8 encodes a myristylated armadillo protein of the *Saccharomyces cerevisiae* vacuolar membrane that functions in vacuole fusion and inheritance.

98158703:  
Vac8p, a vacuolar protein with armadillo repeats, functions in both vacuole inheritance and protein targeting from the cytoplasm to vacuole.

## Peptide information for frame 3

ORF from 99 bp to 1613 bp; peptide length: 505  
Category: similarity to known protein  
Classification: unset

```
1 MVNILDSPHK SLKCLAAETI ANVAKFKRAR RVVRQHGGIT KLVALLDCAH
51 DSTKPAQSSL YEARDVEVAR CGALALWSCS KSHTNKEAIR KAGGIPLLAR
101 LLKTSHEHML IPVVGTLQEC ASEENYRAAI KAERIIEHVL KNLMSNEQL
151 QHCAMAIYQ CAEDKETRDL VRLHGGLKPL ASLLNWDKK ERLAATGAI
201 WKCSISKENV TKFREYKAI E TLVGLLTDOO EEVLVNVVGA LGCECCQEREN
251 RVIVRKCGGI QPLVNLVGI NOALLVNVTK AVGACAVEPE SMHIIDRLDG
301 VRLLSLLKN PHPDVKASAA WALCPCKNA KDAGEMVRSF VGGLELIVNL
351 LKSDNKEVLA SVCAAITNIA KDQENLAVIT DHGVVPLLSK LANTNNKLR
401 HHLAEAISRC CMWGRNRVAF GEKAVAPLV RYLSNDTNV HRATAQALYQ
451 LSEDADNCIT MHENGAVKLL LDMVGSPDQD LQEAAGCIS NIRRLALATE
501 KARYT
```

## BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphtes3\_35p17, frame 3

PIR:S50446 VAC8 protein - yeast (*Saccharomyces cerevisiae*), N = 1,  
Score = 237, P = 7.8e-17

PIR:T00403 T13E15.9 protein - *Arabidopsis thaliana*, N = 1, Score = 215,  
P = 4.9e-14

TREMBL:DR41081.1 product: "b-catenin"; Danio rerio b-catenin mRNA,  
complete cds., N = 1, Score = 195, P = 5.8e-12

>PIR:S50446 VAC8 protein - yeast (*Saccharomyces cerevisiae*)  
Length = 578

## HSPs:

Score = 237 (35.6 bits), Expect = 7.8e-17, P = 7.8e-17  
Identities = 106/401 (26%), Positives = 177/401 (44%)

```
Query: 92 ACGIFLLARLLKTSHEHMLIPVVGTLQECASEENYRAAIKAERIIEHVLKNLMSNEQLQ 151
      +GG FL A +N+ + L E V + E ++E ++ L S++ O+Q
Sbjct: 45 SGG-PLKALTTLVYSDNLNQRSAAALAFAEITEKYVRQVSR-EVLEPILILLQSDPQIQ 102

Query: 152 EHCAMAIYQCAEDKETRDLVRLHGGLKPLASLLNNTDNKERLAATGAIWKCSISKENV 211
      A+ A + E + L+ GGL+PL + + DN E G I + +N
Sbjct: 103 VAACAALGNLAVNNENKLLIVEMGGLEPLINQMGMG-DNVEVQCNAVGCITNLATRDONKH 161

Query: 212 KFREYKAIETLVGLLTDOPEEVLVNVVVGALGCECCQERENRIVRKCGGIQPLVNLVGIN 271
      K A+ L L + V N GAL ENR + G + LV+LL +
Sbjct: 162 KIATSGALIPLTKLAKSKHVRQVQATGALLNTHSEENKELVNAGAVPVLVSLLSSTO 221

Query: 272 QALLVNVTKAVGACAVEPESMHIIDRLDG--VRLLSLLKNPHDPVKASAAWALCPCKIN 329
      + T A+ AV+ + + + + V L SL+ +P VK A AL +
```

Sbjct: 222 PDVQYCTTALSNIADENRKKLAQTEPRLVSKLVSLMDSPPSRVKQATLALRNLASD 281

Query: 330 AKDAGENVRSFVGGLELIVNLLKSDNKE-VLASVCAAITNIAKDQENLAVITDHGVV-PL 387  
E+VR+ GGL +V L++SD+ VLASV A I NI+ N +I D G + PL

Sbjct: 282 TSYQLEIVRA--GGPLHLVKLIQSDSIPLVLASV-ACIRNISIHPLNEGLIVDAGFLKPL 338

Query: 388 LSKLANTNNKLRHHLAEAISRCMMG-RNRVAFGEHKAVAPLVRYLKSNDTNVHRATAQ 446  
+ L ++ ++ H + +NR F E AV +V ++

Sbjct: 339 VRLLDYKDSEIEIQCHAVSTLRNLAASSEKNRKEFFESGAVEKCKELALDSPVSV-QSEIS 397

Query: 447 ALYQLESDAD-NCITMHENGAVKLLDMVGSPPDQLQEAAGCISNI 492  
A + + AD + + + E + L+ M S +Q++ AA ++N+

Sbjct: 398 ACFAILALADVSKLDLEANILDALIPMTFSQNEQVSGHAAALANL 444

Score = 213 (32.0 bits), Expect = 3.6e-14, P = 3.6e-14  
Identities = 81/341 (23%), Positives = 163/341 (47%)

Query: 163 EDKETROLVRLHGGKPLASLLNNTD-NKERLAAVTGAIWKCSISKENVTKFREYKAIET 221  
EDK+ D G LK L +L+ + + N+ R AA+ A I+++ V + + +E

Sbjct: 36 EDKDQDFYS-GGPLKALTTLVYSDNLNQRSAAALFA----EITEKYVRQVSR-EVLEP 89

Query: 222 LVGLLTQDPEEVLVNVVVGALGECCQERENRIVRKCGGIOPVLNVLVGINQALLVNVTKA 281  
++ LL Q ++ V ALG EN+++ + GG++PL+N ++G N + N

Sbjct: 90 ILILLOSQDPQIQVAACAALGNLAVNNENKLLIVEMGGLEPLINQMMDNVEVQCNVAGC 149

Query: 282 VGACAVEPESMNIIIDRLDGVRLWSLLKNPHDPVKASAAWALCPCIKNAKDAGEMVRSFV 341  
+ A ++ I + L L K+ H V+ +A AL + + + E+V + -

Sbjct: 150 ITNLATRDNDKHKIATSGALIPLTKLAKSKHIVQRNATGALLNMTHSEENRKLVA-- 207

Query: 342 GGLELIVNLLKSDNKEVLASVCAAITNIAKDQENLAVI--TDHGVVPLLSKLANTNNK 399  
G + ++V+LL S + +V A++NIA D+ N + T+ +V L L ++ +++

Sbjct: 208 GAVPVLVSLSSDTPDVQYCTTALSNIADENRKKLAQTEPRLVSKLVSLMDSPPSRV 267

Query: 400 RHHLAEAISRCMMGRNRVAFGEHKAVAPLVRYLKSNDTNVHRATAQALYQLESDADNCI 459  
+ A+ + + + + LV+ ++S+ + A+ + +S N

Sbjct: 268 KCQATLALRNLASDTSYQLEIVRAGGLPHLVKLIQSDSIPLVLASVACIRNISIHPLNEG 327

Query: 460 TMHENGAVKLLDMVGSPPDQLQEAAGCISNIRRLALATEKAR 503  
+ + G +K L+ ++ D + E +S +R LA ++EK R

Sbjct: 328 LIVDAGFLKPLVRLLDYKDSE--EIQCHAVSTLRNLAASSEKNR 369

Score = 180 (27.0 bits), Expect = 1.6e-10, P = 1.6e-10  
Identities = 80/346 (23%), Positives = 142/346 (41%)

Query: 145 SENEQLEQHCAMAIYQCAEDKETDRLVRLHGGKPLASLLNNTDNKERLAAVTGAIWKCS 204  
S+N LQ A+A + E K R + R L+P+ LL + D + ++AA A+ +

Sbjct: 58 SDNLNQRSAAALAFABITE-KYVRQVSR--EVLEPILILLOSQDPQIQVAACA-ALGNLA 113

Query: 205 ISKENVTKFREYKAIETLVGLLTQDPEEVLVNVVVGALGECCQERENRIVRKCGGIOPLV 264  
++ EN E +E L+ + + EV N VG + +N+ + G + PL

Sbjct: 114 VNNENKLLIVEMGGLEPLINQMMDNVEVQCNVAGCITNLATRDNDKHKIATSGALIPLT 173

Query: 265 NLLVGINQALLVNVTKAVGACAVEPESMNIIIDRLDGVRLWSLLKNPHDPVKASAAWALC 324  
L + + N T A+ E+ + V+L SLL + PDV+ AL

Sbjct: 174 KLAKSKHIVQRNATGALLNMTHSEENRKLVAAGVPLVSLSSDTPDVQYCTTALS 233

Query: 325 PCIKNAKDAGEMVRSFVGGLELIVNLLKSDNKEVLASVCAAITNIAKDQENLAVITDHGV 384  
+ + ++ ++ + +V+L+ S + V A+ N+A D I G

Sbjct: 234 NIAVDEANRKKLAQTEPRLVSKLVSLMDSPPSRVKQATLALRNLASDTSYQLEIVRAGG 293

Query: 385 VPLLSKLANTNNKLRHHLAEAISRCMMGRNRVAFGEHKAVAPLVRYLKSNDTNVHRAT 444  
+P L KL +++ L I + N + + PLVR L D+ +

Sbjct: 294 LPHLVKLIQSDSIPLVLASVACIRNISIHPLNEGLIVDAGFLKPLVRLLDYKDSEIEIQCH 353

Query: 445 A-QALYQLESDAD-NCITMHENGAVKLLDMVGSPPDQLQEAAGCIS 490  
A L L+ ++ N E+GAV+ ++ +Q + C +

Sbjct: 354 AVSTLRNLAASSEKNRKEFFESGAVEKCKELALDSPVSVQSEISACFA 401

Score = 155 (23.3 bits), Expect = 8.8e-08, P = 8.8e-08  
Identities = 88/401 (21%), Positives = 175/401 (43%)

Query: 60 LYEAR--VEVARCGALALWCSKSHNTKEAIRKAGGI-PLLARLLKTSHENMLIPVVG 116  
L +++D ++VA C AL + + ++ NK I + GG+ PL+ ++ + E + VG

Sbjct: 93 LLQSQDPQIQVAACAALG--NLAVNNENKLLIVEMGGLEPLINQMMDNVE-VQCNVAGC 149

Query: 117 LQECASEENVRAATKAERIENLVKNLSENEQLEQHCAMAIYQCAEDKETR-DLVRLHG 175  
+ A+ ++ + I + L K S++ ++Q + A+ +ER +LV G

Sbjct: 150 ITNLATRDNDKHKIATSGALIPLTKLAKSKHIVQRNATGALLNMTHSEENRKLVA--G 208

Query: 176 GLKPLASLLNNTDNKERLAAVTGAIWKCSISKENVTKFR--EYKAIETLVGLLTQDPEEV 233  
+ L SLL++TD + T A+ ++ + N K E + + LV L+ V

Sbjct: 209 AVPVLSLLSSTDPDVQYCTT-ALSNIAVDENRKKLAQTEPRVSKVSLMDSPPSSRV 267

Query: 234 LVNVVVGALGECQERENRVRKCGGIQPLVNLVGINQALLVNVTKAVGACAVEPESMM 293  
AL + + + + GG+ LV L+ + L++ + + P +

Sbjct: 268 KCQATLALRNLASDTSYQLEIVRAGGLPHLVKLIQSDSIPLVLASVACIRNISIHPLNEG 327

Query: 294 IIDRLDGVRLLSLLK-NPHPDVKASAAWALCPCIKNA-KDAGEMVRSFVGGLELIVNLL 351  
+I ++ L LL ++ A L ++ K+ E S G +E L

Sbjct: 328 LIVDAGFLKPLVRLLDYKDEEIQCHAVSTLRNLAASSEKNRKEFFES--GAVERCKELA 385

Query: 352 KSDNKEVLA--SVCAAITNIAKDQENLAVITDHGVVPLLSKLANTNNKLRHHLAEATSR 409  
V + S C AI +A D L ++ + ++ L + + N ++ + A A++

Sbjct: 386 LDSPVSVQSEISACFAILALA-DVSKLDLL-EANILDALIPMTFSQNEQVSGNAAALAN 443

Query: 410 CCHMGRNRVAFGE----HKAVAP-LVRVYKSNDRVHRATAQALYQISE 453  
C N E ++ + L+R+LKS+ + QL E

Sbjct: 444 LCSRNNYTKIIEAWDRPNEGIRGFLIRFLKSDYATFEHIALWTLQLE 493

Score = 139 (20.9 bits), Expect = 5.0e-06, P = 5.0e-06  
Identities = 80/329 (24%), Positives = 142/329 (43%)

Query: 37 GGITKLVALLOCAHD-STKPAQ---SSLYEARDVEVARCGALALWSCSKSHNTKEAIRKA 92  
G IT L D H + T A L + + + + V R AL + + S N++ + A

Sbjct: 148 GCITHLATRDNDKHKIATSGALIPITKLAKSKHVRQVQATGALLNMTHSENRKELVNA 207

Query: 93 GGIPLLARLLKTSHEMLIPVVGTLQECASEE-NYRAATKAE-RIENLVKNINSENEQL 150  
G +P+L LL ++ + L A +E N + + E R++ LV ++S + +

Sbjct: 208 GAVPVLSLLSSTDPDVQYCTTALSNIADENRKKLAQTEPRVSKVSLMDSPPSSRV 267

Query: 151 QEHCAIYQCAEDKETR-DLVRHGGGLKPLASLLNNTDNKERLAATGAIWKCSISKEN 209  
+ A+ A D + +VR GGL L L+ + D+ + A I SI N

Sbjct: 268 KCQATLALRNLASDTSYQLEIVRA-GGLPHLVKLIQS-DSIPLVLASVACIRNISIHPLN 325

Query: 210 VTKFREYKAIETLVGLLT-DQPEELVNVVVGALGECQERE-NRVRKCGGIQPLVNL 267  
+ + LV LL EE+ + V L E NR + G ++ L

Sbjct: 326 EGLIVDAGFLKPLVRLLDYKDEEIQCHAVSTLRNLAASSEKNRKEFFESGAVERCKELA 385

Query: 268 VG--INQALLVNVTKAVGACA-VEPESMMIIDRLDGVRLLSLLKNPHPDVKASAAW-L 323  
+ + + + A A A V + + LD + + + + N A+AA A L

Sbjct: 386 LDSPVSVQSEISACFAILALADVSKLDLEANILDAL-IPMTFSQNEQVSGNAAALANL 444

Query: 324 CPCIKH-AKDAGEMVRSFVGGLELIVNLLKSD 354  
C + N K R G ++ LKSD

Sbjct: 445 CSRNNYTKIIEAWDRPNEGIRGFLIRFLKSD 476

Score = 136 (20.4 bits), Expect = 1.1e-05, P = 1.1e-05  
Identities = 72/304 (23%), Positives = 133/304 (43%)

Query: 58 SSLYEARDVEVARCGALALWSCSKSHNTKEAIRKAGGIPLARLLKTSHEMLIPVVGTL 117  
+ L + + + + V R AL + + S N++ + AG +P+L LL ++ + L

Sbjct: 173 TKLAKSKHVRQVQATGALLNMTHSENRKELVNAGAVPVLSLLSSTDPDVQYCTTAL 232

Query: 118 QECASEE-NYRAATKAE-RIENLVKNINSENEQLQEHCAIYQCAEDKETR-DLVRH 174  
A +E N + + E R++ LV ++S + + + A+ A D + +VR

Sbjct: 233 SNIAVDENRKKLAQTEPRVSKVSLMDSPPSSRVKCQATLALRNLASDTSYQLEIVRA- 291

Query: 175 GGLKPLASLLNNTDNKERLAATGAIWKCSISKENVTKFREYKAIETLVGLLT-DQPEEV 233  
GGL L L+ + D+ + A I SI N + ++ LV LL EE+

Sbjct: 292 GGLPHLVKLIQS-DSIPLVLASVACIRNISIHPLNEGLIVDAGFLKPLVRLLDYKDEEI 350

Query: 234 LVNVVVGALGECQERE-NRVRKCGGIQPLVNLVVG--INQALLVNVTKAVGACA-VEP 289  
+ V L E NR + G ++ L + + + + A+ A A V

Sbjct: 351 QCHAVSTLRNLAASSEKNRKEFFESGAVERCKELALDSPSVQSEISACFAILADVSK 410

Query: 290 ESMNIIDRLDGVRLLSLLKNPHPDVKASAAW-LCPCIKH-AKDAGEMVRSFVGGLELI 347  
++ + LD + + + + N A+AA A LC + N K R G +

Sbjct: 411 LDLEANILDAL-IPMTFSQNEQVSGNAAALANLCSRNNYTKIIEAWDRPNEGIRGFL 469

Query: 348 VNLKSD 354  
+ LKSD

Sbjct: 470 IRFLKSD 476

Score = 114 (17.1 bits), Expect = 2.7e-03, P = 2.7e-03  
Identities = 71/335 (21%), Positives = 132/335 (39%)

Query: 1 MVNILDSPHKSLKLAETIANVAKFKRARRVROHGGITKLVALLOCAHDSTKPAQSSL 60  
+ + S H ++ A + N+ + R+ + G + LV+LL ST P

Sbjct: 172 LTKLAKSKHVRQVQATGALLNMTHSENRKELVNAGAVPVLSLLS-----STDP----- 222

Query: 61 YEARDVEVARCGALALWSCSKSHNTKEAIRKAGGIPLARLLKTSHEMLIPVVGTLQEC 120  
DV+ AL+ + + + K A + + L L+ + + L+





```
SEQ  ASLLNMTDNKERLAAVTGAIWKCSISKENVTKFREYKAIETLVGLLTDQPEEVLNVVGA
SEG  .....
2bct- HHHHH-HCCCCHHHHHHHHHHHHHHCCCCHHHHHHHHHCHHHHHHTTTTCCHHHHHHHHHH

SEQ  LGECQCQERENRIVRKCGGIQPLVNLVGINQALLVNVTKAVGACAVEPESHMIIDRLDG
SEG  .....
2bct- H-----HHHHHCCCCCTTTTHHHHHHHHHHHHCTTTTHHHHHHHHHHTTNKHHHHHH-HHCH

SEQ  VRLWSSLKNPHPDVKASAAWALCPCKNAKDAGENVRSFVGGLELIVNLLKSDNKEVLA
SEG  .....
2bct- HHHHHHHHTTTTHHHHHHHHHHHHHHHCCCCHH-HHHHHHHHHHHHHHCTTTTHHHH

SEQ  SVCAAITNIAKDQENLAVITDHGVVPLLSKLANTNNKLRHHLAEAISRCCHMGRNRVAF
SEG  .....
2bct- HHHHHHHHHHHC GGHHHHHHHHHHHHHHHHHHHHHTTTTCHHHHHHHHHHHHCHHHHH

SEQ  GENKAVAPLVRYLKSNDTNVHRATAQALYQLSEDAENCITMHENGAVKLLLDWVGSPDOD
SEG  .....
2bct- HTTTTHHHHHHHHCCCCHHHHHHHHHHHHHHHTTTTHHHHHHHHCCCCHHHHHHNTTTTTHH

SEQ  LQEAAGCISNIRRLALATEKARYT
SEG  .....
2bct- HHHHHHHHH.....
```

(No Prosite data available for DKFZphtes3\_35p17.3)

(No Pfam data available for DKFZphtes3\_35p17.3)

DKFZphtes3\_35p22

group: cell cycle

DKFZphtes3\_35p22 encodes a novel 549 amino acid protein, with similarity to oncogene 1 (tre-2 locus).

The novel protein is closely related to human tre-2 and other enzymes involved in the degradation of ubiquitinated proteins. The human tre-2 oncogene encodes a deubiquitinating enzyme, indicating a role for the ubiquitin system in mammalian growth control.

The novel protein can find application in cancer diagnostics and treatment, and in regulating protein stability and growth control via regulation of ubiquitination.

strong similarity to oncogene 1 (tre-2 locus)

membrane regions: 1

complete cDNA, complete cds, EST hits

Sequenced by DKFZ

Locus: map="17"

Insert length: 2072 bp

Poly A stretch at pos. 2062, polyadenylation signal at pos. 2039

```
1 GTTACACACA GGCAGTGGTA TCTGTGAGCA GCTCTGTGGA CTCAAAGGTT
51 TTCTCCCTGA GAGGCATGAC CCAGGCCAGC TGATTCATCA GAATCAGGAT
101 GGACGTGGTA GAGGTCCGGG GCAGTTGGTG GGCACAAGAG CGAGAGGACA
151 TCATTATGAA ATACGAAAAG GGACACCGAG CTGGGCTGCC AGAGGACAAAG
201 GGGCCTAAGC CTTTTCGAAG CTACACAAAC AACCTCGATC ATTTGGGGAT
251 TGTACATGAG ACGGAGCTGC CTCCTCTGAC TGGCGGGGAG GCGAAGCAAA
301 TTGGCGGGGA GATCAGCCGA AAGAGCAAGT GGGTGGATAT GCTGGGAGAC
351 TGGGAGAAAT ACAAAAGCAG CAGAAAGCTC ATAGATCGAG CGTACAAGGG
401 AATGCCCATG AACATCCGGG GCCCGATGTG GTCAGTCCCT CTGAACACTG
451 AGGAAATGAA GTTGA AAAAC CCCGGAAGAT ACCAGATCAT GAAGGAGAG
501 GGCAGAAGT CATCTGAGCA CATCCAGCGC ATGACCGGG ACCTAAGCGG
551 GACATTAAAG AAGCATATAT TCTTCAGGCA TCGATACGGA ACCAAGCAGC
601 GGGAACTACT CCACATCCTC CTGGCATATG AGGAGTACAA CCGGAGGTG
651 GGCTACTGCA GGGACCTGAG CCACATCGCC GCCTTGTTC TCTCTATCT
701 TCTGAGGAG GATGCATTCT GGGCACTGGT GCAGCTGCTG GCCAGTGAGA
751 GGCATCCCT GCAGGGATT CACAGCCCAA ATGGCGGGAC GCTCAGGGG
801 CTCGAGACC AACAGGACCA TGTGTTAGCC ACCTCACAAC CCAAGACCAT
851 GGGCATCTAG GACAAGAAAG ATCTATGTGG GCAGTGTTC CCGTTAGGCT
901 GCCTCATCCG GATATTGATT GACGGGATCT CTCCTGGGCT CACCTGCGC
951 CTGTGGGACG TGTATCTGTT AGAAGGCGAA CAGGCGCTGA TGCCGATAAC
1001 AAGAAATGCC TTTAAGGTTT AGCAGAAGCG CCTCAGGAAG ACCTCCAGGT
1051 GTGGCCCGTG GGCACGTTTT TGCAACCGGT TCGTTGATAC CTGGGCCAGG
1101 GATGAGGACA CTGTGCTCAA GCATCTTAGG GCCTCATGGA AGAATTAAC
1151 AAGAAAGAG GGGGACCTGC CACCCCGCAG CAAACCCGAG CAAGGGCTGT
1201 GGGCATCCAG GCCTGTGCGG GCTTCAGTG GCGGGAAGAC CCTCTGCAAG
1251 GGGGACAGGC AGGCCCTCC AGGCCACCA GCCCGGTTCC CGCGGCCCAT
1301 TTGGTCAGCT TCCCGGCCAC GGGCACCTCG TTCTTCCACA CCCTGTCTGT
1351 GTGGGCTGT CCGGGAAGAC ACCTACCTTG TGGGCACTCA GGGTGTGCC
1401 AGCCCGGCC TGGCTCAGGG AGGACTCAG GGTTCCTGGA GATTCTGCA
1451 GTGGAACCTC ATGCCCGCC TCCAAAGGGA CCTGGAGCTA GAGGGCCCTT
1501 GGTTCGCCCA TTATGATTTC AGACAGAGCT GCTGGGCTCC TGCCATATCC
1551 CAGGAGGACC AGCTGGCCCC CTGCTGGCAG GCTGAACACC CTGCGGAGCG
1601 GGTGAGATCG GCTTTCGCTG CACCCAGCAC TGATTCCGAC CAGGGCACCC
1651 CTTTCAAGAG TAGGACGAGA CAGCAGTGTG CTCCACCTC AGGGCCTTGG
1701 CTCTGGGCC TCCACTTGGG AAGTTCTCAG TTCCCTCAG GATTCTGAA
1751 GCATCTGGGC CAGGGCTCAT GCGTGATAA TTCCCTAGG CTTAACAACC
1801 CAGCAAGCT TCGCATCCTC GTTTTATTT TGGTTAAACT TATGAAAATG
1851 TATTAAGAAA GAGTGCAGCT CGAGAGAGAT TCAGAGATGG AACACACCAG
1901 ACCCAGATC ACAAGGCCAA CCATGCCAG CCCCTCCAG CACCCACAGC
1951 CCCACGACA TCGTTCTGAA TTCTGACGAC ACCGTGAGCC TGCTTTGTA
2001 CTCAAACTC ATGGAAGGAT AACCACTTC ATGTTTTGAA ATAAATGTTT
2051 CCGTTGAAA TGAAAAAAA AA
```

## BLAST Results

Entry AC003976 from database EMBL:  
Homo sapiens chromosome 17, clone hCIT.91 J.4, complete sequence.  
Score = 4385, P = 0.0e+00, identities = 8817886

## 14 exons

Entry HSG19723 from database EMBL:  
human STS A001W35.  
Score = 850, P = 1.9e-32, identities = 170/170

## Medline entries

92228503:  
A novel transcriptional unit of the tre oncogene widely  
expressed in human cancer cells.

94067315:  
The yeast DOA4 gene encodes a deubiquitinating enzyme  
related to a product of the human tre-2 oncogene.

95176708:  
UBP5 encodes a putative yeast ubiquitin-specific protease  
that is related to the human Tre-2 oncogene product.

## Peptide information for frame 3

ORF from 99 bp to 1745 bp; peptide length: 549  
Category: strong similarity to known protein

```
1 MDVVEVAGSW WAQEREDIIM KYEKGHRAGL PEDKGPKPFR SYNNVVDHLG
51 IVHETELPPL TAREAKQIRR EISRKSKWVD MLGOMEXYKS SRKLIDRAYK
101 QMPNMIRGPM WSVLLNTEEM KLNKNGRYQI MKEGKKSSE HIQRIQDVVS
151 GTLRKHIFER DRYGTQKREL LHILLAYEY NPEVGRCROL SHIAALFLLY
201 LPEEDAFWAL VOLLASERHS LQGFHSPNGG TVQGLQDQOE HVVATSPKPT
251 MGHQDKKDLG GOCSPGLCLI RILIDIGISLG LTLRLWDVYL VEGEQALMPI
301 TRIAFKVOQR RLTKTSRCGP WAFPCNRFDV TWARDDETVL KHLRASHKSL
351 TRAKGGLPFP ARFEQSSAS RPVPASRGCK TLCKGQKQAP PQPPAREFRP
401 IWSASPPRAP RSTPCPCGA VREDTYPVCT QGVSPALAQ GQPGQSHRFL
451 QWNSMPLRPT DLVDEGPNFR HYDFRQSCWV RAISQEDOLA PCWQAEHPAE
501 RVRSAFAAPS TDSQDGTFFR ARDEQQAAPT SGPCLCGLHL ESSQFPFPGF
```

## BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphtes3\_35p22, frame 3

PIR:S22155 oncogene 1 (tre-2 locus) (clone 210) - human, N = 1, Score =  
2181, P = 5.5e-226

PIR:S57867 oncogene 1 - human, N = 1, Score = 1536, P = 1.2e-157

>PIR:S22155 oncogene 1 (tre-2 locus) (clone 210) - human  
Length = 786

## HSPs:

Score = 2181 (327.2 bits), Expect = 5.5e-226, P = 5.5e-226  
Identities = 405/500 (81%), Positives = 440/500 (88%)

```
Query: 1 MDVVEVAGSWWAQEREDIIMKYEKGHRAGLPEDKGPKPFRSYNNVVDHLGIVHETELPPL 60
MD+VE A S AQER+DI+MKY+KGHRAGLPEDKG+P N+++D GI+HETELPP+
Sbjct: 1 MDNVENADSLQAQERKIDILMKYDKGHRAGLPEDKGPEPV-GINSSIDRFGILHETELPPV 59

Query: 61 TAREAKQIRREISRKSKWVMDLGOMEXYKSSRKLIDRAYKQMPNMIRGPMWSVLLNTEEM 120
TAREAK+IRRE++R SKW++MLG+WE YK S KLIDR YKG+PMNIRGP+WSVLLN +E+
Sbjct: 60 TAREAKKIRREMTRTSKWMEMLGEWETKYKSSKLIDRVYKIPMNIRGPMVSVLLNIQEI 119

Query: 121 KLNKNGRYQIMKEGKKSSEHIQRIQDVSGTLRKHIFRDYGTQKRELLHILLAYEY 180
KLNKNGRYQIMKE+GK+SSEHI ID DV TLR H+FRDRYG KQREL +ILLAY EY
Sbjct: 120 KLNKNGRYQIMKEGKKSSEHIHIDLDVTRTLRNHVFRDRYGAKQRELFYILLAYEY 179

Query: 181 NPEVGRCRDLSHIAALFLLYLPEEDAFWALVOLLASERHSLQGFHSPNGGTVQGLQDQOE 240
NPEVGRCRDLSHI ALFLLYLPEEDAFWALVOLLASERHSL GFHSPNGGTVQGLQDQOE
Sbjct: 180 NPEVGRCRDLSHITALFLLYLPEEDAFWALVOLLASERHSLPGFHSPNGGTVQGLQDQOE 239
```

Query: 241 HVVATSQPKTMGHQDKDLCCGCSPLGCLIRILIDGISLGLTLRLWDVYLVEGEQALMPI 300  
HVV SQPKTM HQDK+ LCGC+ LGCL+R LIDGISLGLTLRLWDVYLVEGEQ LMPI  
Sbjct: 240 HVVPKSQPKTMGHQDKGELCGCASLGCLLRNLIDGISLGLTLRLWDVYLVEGEQVLMPI 299

Query: 301 TRIAFKVOQKRLTKTSRCGPWAFRCNRFVDTWARDVTLKHLRASMKLTTRKKGDLPPP 360  
T IA KVQKKRL KTSRCG WAR N+P DFWA ++DTVLKHLRAS KLTTR+GDLPPP  
Sbjct: 300 TSIALKVQKKRLTKTSRCGLMARLNRQFFDTWAMNDVTLKHLRASKLTTRKQGDLP 359

Query: 361 AKPEQSSASRPVPASRGKTLCKGDRQAPPGPPARFPPIWASPPRAPRSSTPCPGGA 420  
AK EQGS A RPVPASRGKTLCKG RQAPPGPPA+P RPI SASPP A R STPCPGGA  
Sbjct: 360 AKREQGS LAPRPVPASRGKTLCKGYRQAPPGPPAQFQRPICSAFPWASRFSTPCPGGA 419

Query: 421 VREDTYPVGTQGVPSALAQGGPOGWSRFLWNSMPRLPTDLDVEGFWFRHYDFRQSCWV 480  
VREDTYPVGTQGVPS ALAQGGPOGWSRFL+W SMPRLPTDLD+ GPF HYDF +SCWV  
Sbjct: 420 VREDTYPVGTQGVPSALAQGGPOGWSRFLWNSMPRLPTDLDIGGFWFRHYDFRSCWV 479

Query: 481 RAISQEDQLAPCWAQAEHPAE 500  
RAISQEDQLA CWAQAEH E  
Sbjct: 480 RAISQEDQLATCWAQAEHCGE 499

Pedant information for DKFZphtes3\_35p22, frame 3

Report for DKFZphtes3\_35p22.3

[LENGTH] 549  
[MW] 62159.16  
[PI] 9.23  
[HOMOL] PIR:S22155 oncogene 1 (tre-2 locus) (clone 210) - human 0.0  
[FUNCAT] 11.01 stress response [S. cerevisiae, YGR100w] 2e-16  
[FUNCAT] 04.05.01.04 transcriptional control [S. cerevisiae, YGR100w] 2e-16  
[FUNCAT] 99 unclassified proteins [S. cerevisiae, YNL293w] 3e-15  
[PIRKM] transmembrane protein 6e-14  
[PROSITE] MYRISTYL 6  
[PROSITE] AMIDATION 1  
[PROSITE] CAMP\_PHOSPHO\_SITE 3  
[PROSITE] CK2\_PHOSPHO\_SITE 4  
[PROSITE] TYR\_PHOSPHO\_SITE 2  
[PROSITE] PKC\_PHOSPHO\_SITE 10  
[KW] TRANSMEMBRANE 1  
[KW] LOW\_COMPLEXITY 5.28 %

SEQ MDVVEVAGSMAQEREDIIMKYEGHRAGLPEDKGPFPFRSYNNVVDHLGIVHETELPPL  
SEG .....  
PRD cccccccccchhhhhhhhhhhhhhhcccccccccccccccccccccccccccccccccccc  
MEM .....  
SEQ TAREAKIRREISRSKSWVDMLGDEWKYKSRKIDRAYKGMPMNIAGPMMSVLLNTEEM  
SEG .....  
PRD chhhhhhhhhhhhhhhhhhhhhhhcccccccccccccccccccccccccccccccccccc  
MEM .....  
SEQ KLKNPGRYQIMKEGKKSSEIQRIDROVSGTLRKHIFFRDYGTKQRELLHILLAYEY  
SEG .....  
PRD cccccccccchhhhhhhhhhhhhhhhhhhhhhhcccccccccccccccccccccccccccc  
MEM .....  
SEQ NPEVGYCDLSHIAALFLYLPEEDAFWALVQLASERHSLQGFHSPNGGTVOGLQDOQE  
SEG .....  
PRD cccccccccchhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhcccccccccccccccccc  
MEM .....  
SEQ HVVATSQPKTMGHQDKDLCCGCSPLGCLIRILIDGISLGLTLRLWDVYLVEGEQALMPI  
SEG .....  
PRD hhhhhhhchhhhhhhhhhhhhhhccccccccchhhhhhhhhhhhhhhccccccccchhhhh  
MEM .....  
SEQ TRIAFKVOQKRLTKTSRCGPWAFRCNRFVDTWARDVTLKHLRASMKLTTRKKGDLPPP  
SEG .....  
PRD hhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhccccccccchhhhhhhhhhhhhhhhhccc  
MEM .....  
SEQ AKPEQSSASRPVPASRGKTLCKGDRQAPPGPPARFPPIWASPPRAPRSSTPCPGGA  
SEG .....  
PRD ccc  
MEM .....

WO 01/12659

PCT/IB00/01496

```
SEQ VREDTYPVGTQGVSPALAQGGPQGSWRFLQNSMRLPTDLDVEGPFWRHYDFRQSCWV
SEG .....
PRD ccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc
MEM .....

SEQ RAISQEDQLAPCWAEPFAERVRSAFAAPSTDSQGTFFRAREQQCAPTSGLCLGLHL
SEG .....
PRD cchhhhhhhhhhhhhcchhhhhhhccccccccccccchhhhhcccccccccccccccc
MEM .....

SEQ ESSQFPFGF
SEG .....
PRD ccccccccc
MEM .....
```

Prosite for DKF2phtes3\_35p22.3

PS00004	136->140	CAMP_PHOSPHO_SITE	PDOC00004
PS00004	310->314	CAMP_PHOSPHO_SITE	PDOC00004
PS00004	348->352	CAMP_PHOSPHO_SITE	PDOC00004
PS00005	61->64	PKC_PHOSPHO_SITE	PDOC00005
PS00005	73->76	PKC_PHOSPHO_SITE	PDOC00005
PS00005	90->93	PKC_PHOSPHO_SITE	PDOC00005
PS00005	152->155	PKC_PHOSPHO_SITE	PDOC00005
PS00005	216->219	PKC_PHOSPHO_SITE	PDOC00005
PS00005	282->285	PKC_PHOSPHO_SITE	PDOC00005
PS00005	315->318	PKC_PHOSPHO_SITE	PDOC00005
PS00005	346->349	PKC_PHOSPHO_SITE	PDOC00005
PS00005	351->354	PKC_PHOSPHO_SITE	PDOC00005
PS00005	446->449	PKC_PHOSPHO_SITE	PDOC00005
PS00006	61->65	CK2_PHOSPHO_SITE	PDOC00006
PS00006	460->464	CK2_PHOSPHO_SITE	PDOC00006
PS00006	484->488	CK2_PHOSPHO_SITE	PDOC00006
PS00006	511->515	CK2_PHOSPHO_SITE	PDOC00006
PS00007	93->100	TYR_PHOSPHO_SITE	PDOC00007
PS00007	92->100	TYR_PHOSPHO_SITE	PDOC00007
PS00008	8->14	MYRISTYL	PDOC00008
PS00008	101->107	MYRISTYL	PDOC00008
PS00008	230->236	MYRISTYL	PDOC00008
PS00008	276->282	MYRISTYL	PDOC00008
PS00008	366->372	MYRISTYL	PDOC00008
PS00008	441->447	MYRISTYL	PDOC00008
PS00009	134->138	AMIDATION	PDOC00009

(No Pfam data available for DKF2phtes3\_35p22.3)

DKFZphtes3\_4b4

group: testes derived

DKFZphtes3\_4b4 encodes a novel 497 amino acid protein similar to SCP proteins and a human trypsin inhibitor.

The novel protein contains an extracellular proteins SCP/Tpx-1/Ag5/PR-1/Sc7 signature 2, predicted by Prosite and Pfam. This domain is found in a variety of extracellular proteins from eukaryotes that have been found to be evolutionary related. The exact function of these proteins is not yet known. In addition, the protein is similar to a human trypsin inhibitor.

No informative BLAST results: No predictive prosite, pfam or SCOP motif.

The new protein can find application in studying the expression profile of testis-specific genes or as a new protease inhibitor.

strong similarity to trypsin inhibitor

might be a new protease inhibitor?

Sequenced by AGOWA

Locus: /map="333.4 cR from top of Chr16 linkage group"

Insert length: 4574 bp

Poly A stretch at pos. 4551, polyadenylation signal at pos. 4539

```
1 GCGGGCTGCT CCCATTGAGC TGTCTGCTCG CTGTGCCCGC TGTGCTGCT
51 GTGCGCCGCG GTGCGCCGCT GCTACCGGGT CTGTGGGACG CGGGAGAGCG
101 CAGCGAGCTG GTGATTGGAG CCTGCGGGAG AGCTCAACCG CCCAGCTCTG
151 CCCGAGGAGC CCAGGCTGCC CCGTGAGTCC CATAGTTGCT GCAGGAGTGG
201 AGCAGTAGCG TGCCTCTGCG GTGGTGCTAT CCCCTGGGGG CTGCTGTTCG
251 TGGTCTGCGG ATCCCAAGGC TACTCTCTGC CCAAGCTCAC TCTCTTAGAG
301 GAGCTGCTCA GCAAAATACCA GCACAACGAG TCTCATCCCC GGGTCCGCG
351 AGCCATCCCC AGGGAGGACA AGGAGGAGAT CCTCATGCTG CACAACAAGC
401 TTCGGGGCCA GGTGCAGCCT CAGGCCCTCCA ACATGGAGTA CATGACCTGG
451 GATGACGAAC TGGAGAAGTC TGCTCAGCG TGCGCAGTTC AGTGCATCTG
501 GAGCAGCGG CCCACCAATC TGTCTGTCTC CATCGGGGAG AACCTGGGCG
551 CTCACTGGGG CAGGTATCGC TCTCCGGGGT TCCATGTGCA GTCTGTGTAT
601 GACGAGGTGA AGGACTACAC CTACCCCTAC CCGAGCGAGT GCAACCCCTG
651 GTGTCCAGAG AGGTGCTCGG GGCCTATGTG CACGCACCTAC ACACAGATAG
701 TTTGGGCCAC CACCAACAAG ATCGGTTGTG CTGTGAACAC CTGCCGGAAG
751 ATGACTGTCT GGGGAGAAGT TTGGGAGAAG CGGCTCTACT TTGTCTGCA
801 TTAATCTCCA AAGGGAACTT GGATTGGAGA AGCCCCCTAC AAGAATGGCC
851 GGCCTGTCTC TGAGTGCCCA CCCAGGTATG GAGGCAGCTG CAGGAACAAC
901 TTGTGTTACC GAGAAGAAAC CTACACTCCA AAACCTGAAA CGGACGAGAT
951 GAATGAGGTG GAAACGGCTC CATTCTCTGA AGAAAACCAT GTTTGCTCTC
1001 AACCGAGGCT GATGAGACCC ACCAAGGCCA AGAAACCTC TGGGTCAAC
1051 TACATGACCC AAGTCGTGAG ATGTGACACC AAGTAGAAGG ACAGGTGCAA
1101 AGGGTCCACG TGTAACAGGT ACCAGTGCCC AGCAGGCTGC CTGAACCACA
1151 AGGCGAAGAT CTTTGGAACT CTGTTCTATG AAAGCTCGTC TAGCATATGC
1201 GCGCGGCCCA TCCACTACGG GATCTCTGAT GACAAGGGAG GCTGTGTGGA
1251 TATCACCAAG AACCGGAAGG TCCCTCTCTT COTGAAGCTC GAGAGACAGC
1301 GCGTCAGTCT CCTCAGCAAA TACAACCTTT CCAGCTCATT CATGGTGTCA
1351 AAAGTGAAAG TGCAGGATTT GGACTGCTAC ACGACCGTTG CTCAGTGTGT
1401 CCGGTTTGAA AAGCCAGCAA CTCACTGCCC AAGAATCCAT TGTCCGGCAC
1451 ACTGCAAGAA CGAACCTTCC TACTGGGCTC CGGTGTTTGG AACCAACATC
1501 TATCGAGATA CCTCAAGCAT CTGCAAGACA GCGGTGCACG CGGAGTCAAT
1551 CAGCAACGAG AGTGGGGGTG ACGTGGACGT GATCGCCGTG GATAAAAGAA
1601 AGACCTACGT GGGCTCGCTC AGGAATGGAG TTCAGTCTGA AAGCCTGGGG
1651 ACTCCTCGGG ATGGAAAGGC CTTCGGGATC TTTGCTGTCA GGCAGTGAAT
1701 TTCCAGCACC AGGGGAGAAG GGGGCTCTTC AGGAGGGCTT CGGGGTTTTG
1751 CTTTATTTT TATTTGTCTA TTGCGGGGTA TATGGAGAGT CAGGAACCTT
1801 CTTTGACTG ATGTTCACTG TCCATCACTT TGTGGCCTGT GGTGAGGTG
1851 ACATCTCATC CCCTCACTGA AGCAACAGCA TCCCAAGGTG CTCAGCCGGA
1901 TCCCTGGTGC CCTGATCCTG CTGGGGCCCG GGGGTCTCCA TCTGACCTG
1951 CTCTCTCCTT TAGAGATCTG AGCTGTCTCT TAAAGGGGAC AGTTGCCCAA
2001 AATGTCTCTT GCTATGTGTT CTCTGTGTGG TGGAGGAAGT TGATTTCAAC
2051 CTCTCTGCTA AAAGAACAAA CCATTTGAGG CTCACAATTG TGAAGCATTC
2101 ACGGCGTCTG AACAGGCTTT TTGAGCAAGC GCCAATGAGT TTCAGGAATG
2151 AAGTAGAAGG TAGTTATTTA AAAATAAAAA ACACAGTCCG TCCCTACCAA
2201 TAGAGGAAAA TGGTTTAAAT GTTTGCTGGT CAGACAGACA AATGGGCTAG
2251 AGTAAGAGGG CTGCGGGTAT CAGAGACCCC GGGTCCGCCC TGGCACGTGT
2301 CTTTCTGTGC GGGCGGCAC AGGCCCTCTT CAATGGCGCG ATTCAAGATG
2351 GCTCTATACA CAGCACTGCT GTTTATGTGA GAGTTACAGA GTCACTTCAG
2401 AGATGTATCT TGTCTTTGTC AGGCCCTTCA TCTTATGCG CCACCTGTTT
2451 TCTGCCGTGA CCTTTGGTCC CATTGAGGAC TAAGATCGG GACCCTTTCT
```

```
2501 TTACCCCTTA CCGATTGTGG CTCCACCCCT GCCTCGGACT GGTTCACGTG
2551 TCCTGGTTCA CACCAGGAC TTTCTTTTC AAGCGAACCT GTTGAAGCC
2601 CAAGTCTTAA CTCCTGGTCT CGTAAGGTTT CACTGAGACG AGATGTCTGA
2651 GAACAACCAA AGAAGGCGCTG CTCCTTGTCTG CTTTAAAAA ATGACAATTA
2701 AATGTGCAGA TTCCCCAGCG ACCCGATGAC CTAATTTTC AGCGTGGA
2751 GGAATGGAGT CTTTGTGACA TTCCCTACCG AGGTTAGCAG CTCAGTTTGT
2801 GGTATGAAA CCGTCTGTGG CCTCATGACA GCGAGAGATG GGAATACACT
2851 AGAAGGATCT CTTTCTCTGT TTTCTGAAA GCAGCTTTGC CAAACGTTCC
2901 CGAGGCGCCA AGGAGTGTAG TACACCTGG CTGCCATCAC TCTATAAAG
2951 TGCTTCATGA GCCCAGACCA AAGCCACACA GTGAATGAA GTACCTTTT
3001 GTAATAGCA TTTTGTGCA GAAGGTGAAA ATTCACACT CTACCACGG
3051 GCCAGCCAAT AGATCACTTT GGTGAATGCT AGTTTCAAA TTGATTCAAA
3101 ATATTCTTA GTGAAAGAA CTAGCAGAAA GTCAAAAAC TAAATACGTG
3151 AGACTGGACA AGAATTCTA CCGGGGACC TAGTGATGC CTCTTTCTT
3201 TGATTGCCTT TCTAATAAT GCAGAACTG AAGTAAATA GGTTTAAAC
3251 AAACAATAA CCAACCCCTT TAAGGATGTG GTAAAAGCA GTTCACTCT
3301 TAGCTTGACT GAGCTAAAAT TCACAGGACT AGTGCTTTG TGCATTGTAG
3351 TCTAGTCGA ATTCAAGGT ACTGACTCTT CAGCCCAAA TGTGGAGAG
3401 GAAGAATTCG GTCAGCCTGT CAGGTCTGTA GTCCAGTTAC CACCAACAT
3451 CTGGGAACCT TCTGGTGTCT GGTGTCTCTG CTGCTGACT TTGTGGCTG
3501 TGCTGTGTCT TGCAAGATAA ATTAGATGCG CCGTGCGGT TTGCAGATT
3551 AGTGAAGGCT CCAGGACGAT GCGAGTGGG TCCTTCCAA AGCATCCAC
3601 TCAAGGAGA CTTGAAACT CAGGTGTGAG TTGACCCAT CATTTAAAA
3651 TAAAGTCCCC GGGTTCTTA ATGCTCTCTT CACTGGGCTT TCTAGCAGG
3701 ATAGAAAGTC CTTGCCAGA GCAGGACCTG GCTGTCTTTT TTTTTTTTT
3751 TTCCCGAGA CCAAGTTTCA CTCGTGTGCG CAGGTAGAG TGCAGTGGG
3801 TGATCTCTCG TCATTGCAAC TGCCGCTCTC CGGTTCAAG CAATCTCAT
3851 GCATCAGCCT CCCAATACC TGGGATACA GCGCTGAGT ACCATGCCC
3901 GCTAATTTTT GTATTTTTAG TAGAGATGGG GTTTCATTAT GTTGGCCAGG
3951 CTGGTCTCGA ACTCCTTACC TCAGGTGATC CACCCACCTT GGCCTCCGA
4001 AGTGTGGGA TTACAGCAT GAGCACTGC GCCCGCCAT GACTGTGCT
4051 GTCCTTATCA TCCCCACAAA CATTGTGAAA CTGGAATATT TGTCTTACA
4101 AAATGGAAC AAGACTATA ATGATAAGCC CTGCTCCAT CACCACCTCT
4151 CCTGTGTGTG GAATAGAGGC CCTCGTGTCT ACCAACACTT ACCCTGTGTT
4201 TAAAAAGATC TTGTACCAAG CCAACGCGGT TCCTGGCTCT CCGCCACAA
4251 GGATGAACAT TTTCGGCTTC CTAGGAGTT TTGCCCTACC GTATTCGAAA
4301 GCGTGTCTGT GTTCTGATA TTGTGTGAG GCTCACTAG CCGCAGTTT
4351 ATGTGTGTCT TTTTCTTAT GAAAAATCAT GTATTTGCT ACTTCTGTG
4401 TACAAAGTTT TATTGTAAAT GTTTTTGTG CTTGATGA ACAGGGGCA
4451 CTTTGTGCA ATTGTTTCAG TAGAAGTGT TTGATTTCTA AATGTTCTT
4501 GTACATATC TTTTATGAAC AAATCTGAAC AATTGTGAA ATAAACACTT
4551 GAAACCAA AAAAAAAA AAAA
```

## BLAST Results

Entry HSB34352 from database EMBL:  
human STS WI-15502.  
Score = 1331, P = 5.4e-54, Identities = 287/301

## Medline entries

98146272:  
cDNA cloning of a novel trypsin inhibitor with similarity to  
pathogenesis-related proteins, and its  
frequent expression in human brain cancer cells.

## Peptide information for frame 1

ORF from 205 bp to 1695 bp: peptide length: 497  
Category: strong similarity to known protein

```
1 MSCVLGGVIP LGLLFLVCGS QGVLLPNVTL LEELLSKYQH NESHRSVRRA
51 IPREDKEEIL MLHNKLAGOV QPOASNMEYM TWDDELEKSA AAWASQIWE
101 HGPTSLVSI GQNLGAHNGR YRSPGFHVQS WYDEVKDYTY PYPSECNPC
151 PERCSGPMCT HYTOIWMATT NKIGCAVNTC RMTVMGEVM ENAVYFVCHY
201 SPKGNWIGEA PYKNGRPCSE CFFSYGGSCR NNLCYREETY TPKPETDEHN
251 EVETAPIEE NHWLPQRMV RPTKPKTSA VNYVQVQVC DTKMDKDCIG
301 STCNRYQCPA GCLNHHKIF GTLFYESSSS ICRAAIHYGI LDDKGLVDI
351 TRNGRVFFV KSERHGVQSL SKYKPSSEFM VSKVKVQDL CYTTVAQLCP
401 FEKPATHCPR IHCPAHKDE PSYWPVFCT NIYADTSSIC KTAHVAGVIS
```



451 NESGGDVDM PVOKKTYVG SLRNGVQSES LGTPRDGKAF RIFAVRQ

## BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphtes3\_4b4, frame 1

TREMBLNEW:AF109674.1 gene: "Lgll"; product: "late gestation lung protein 1"; Rattus norvegicus late gestation lung protein 1 (Lgll) mRNA, complete cds., N = 1, Score = 968, P = 1.9e-97

TREMBL:D45027.1 product: "25 kDa trypsin inhibitor"; Homo sapiens mRNA for 25 kDa trypsin inhibitor, complete cds., N = 1, Score = 738, P = 4.5e-73

TREMBL:AB009609.1 gene: "HrTT-1"; Halocynthia roretzi HrTT-1 mRNA, complete cds., N = 1, Score = 345, P = 2e-31

PIR:JC5308 testis-specific, vespid, and pathogenesis-related protein 1 precursor - human, N = 1, Score = 337, P = 1.7e-30

>TREMBLNEW:AF109674.1 gene: "Lgll"; product: "late gestation lung protein 1"; Rattus norvegicus late gestation lung protein 1 (Lgll) mRNA, complete cds.

Length = 188

## HSPs:

Score = 968 (145.2 bits), Expect = 1.9e-97, P = 1.9e-97  
Identities = 160/185 (86%), Positives = 170/185 (91%)

Query: 61 MLHNKLRGOVOPQASNEHYMTWDDLEKSAWAASQCIWENGPTSLVLSIGQNLCAHMG 120  
MLHNKLRGOV P ASNNEYMTWD+ELE+SAANA +C+WEHGP SLLVLSIGQNL HMGR  
Sbjct: 1 MLHNKLRGOVYPPASNNEYMTWDEELERSAAWAQRCLEWENGPSLLVLSIGQLAVHMG 60

Query: 121 YRSPGFHVQSWYDEVKDYTYPPSECNWPCPERCSGPMCTHTYQIWMATTNKIGCAVNTC 180  
YRSPGFHVQSWYDEVKDYTYPP ECNPWPCPERCSG MCTHTYQ+VMATTNKIGCAV+TC  
Sbjct: 61 YRSPGFHVQSWYDEVKDYTYPPSECNWPCPERCSGAMCTHTYQHWATTNKIGCAVHTC 120

Query: 181 RKMTVWGEVWENAVYVFCNYSKGNWIGEAPYKNGRPCSECPSPSYGGSCRNNLCYREETY 240  
R M+VMG++WENAVY VCNYSKGNWIGEAPYK+GRPCSECP SYGG CRNNLCYREE Y  
Sbjct: 121 RSMVWGDWENAVYVLCNYSKGNWIGEAPYKHGRPCSECPSSYGGSCRNNLCYREEHY 180

Query: 241 TPKEPE 245  
KPE  
Sbjct: 181 HQKPE 185

Pedant information for DKFZphtes3\_4b4, frame 1

## Report for DKFZphtes3\_4b4.1

[LENGTH] 497  
[MW] 55920.00  
[pI] 8.36  
[HOMOL] TREMBL:D45027.1 product: "25 kDa trypsin inhibitor"; Homo sapiens mRNA for 25 kDa trypsin inhibitor, complete cds. 6e-78  
[FUNCAT] 99 unclassified proteins [S. cerevisiae, YJL078c] 8e-12  
[BLOCKS] BLO1009E Extracellular proteins SCP/Tpx-1/Ag5/PR-1/Sc7 proteins  
[BLOCKS] BLO1009D Extracellular proteins SCP/Tpx-1/Ag5/PR-1/Sc7 proteins  
[BLOCKS] BLO1009C Extracellular proteins SCP/Tpx-1/Ag5/PR-1/Sc7 proteins  
[BLOCKS] BLO1009A Extracellular proteins SCP/Tpx-1/Ag5/PR-1/Sc7 proteins  
[PIRKM] glycoprotein 5e-22  
[PIRKM] blocked amino end 5e-13  
[PIRKM] brain 9e-30  
[PIRKM] hydrolase 4e-09  
[PIRKM] hemolymph coagulation 4e-09  
[PIRKM] zymogen 4e-09  
[PIRKM] alternative splicing 4e-09  
[PIRKM] sperm 5e-22  
[PIRKM] viroid-induced protein 2e-11  
[PIRKM] venom 6e-18  
[PIRKM] pyroglutamic acid 2e-11  
[PIRKM] transmembrane protein 2e-10  
[PIRKM] serine proteinase 4e-09  
[SUPFAM] C-type lectin homology 4e-09  
[SUPFAM] trypsin homology 4e-09

[SUPFAM] complement factor H repeat homology 4e-09  
[SUPFAM] cysteine-rich secretory protein 1 6e-24  
[SUPFAM] pathogenesis-related leaf protein 7e-15  
[PROSITE] MYRISTYL 8  
[PROSITE] CAMP\_PHOSPHO\_SITE 3  
[PROSITE] CK2\_PHOSPHO\_SITE 6  
[PROSITE] TYR\_PHOSPHO\_SITE 1  
[PROSITE] PKC\_PHOSPHO\_SITE 8  
[PROSITE] ASN\_GLYCOSYLATION 3  
[PROSITE] SCP\_AG5\_PRI\_SC7\_2 1  
[PFAM] SCP-like extracellular Proteins  
[KW] All\_Beta  
[KW] SIGNAL PEPTIDE 23  
[KW] LOW\_COMPLEXITY 1.21 4

SEQ MSCVLGGVPLGLLFLVCGSQGYLLPNVTLLELLSKYQHNEHSRVRRAIPREDKEIL  
SEG .....KXXXXX.....  
PRD ccc

SEQ MLHNKLKRGQVQPGASNEYMTWDOLEKSAAWASQCIMENGPTSLVSIQNLGAHNGR  
SEG hhhhhhhcc  
PRD ccc

SEQ YRSFGHVQSWYDEVKDTYTPYSECNPMWPERCSGPMCTHYTQIVWATNKGCAVNTC  
SEG .....  
PRD ccc

SEQ RKMTVMGEVWENAVYFVCNYSKGNWIGEPYKNGRPCSECPSPSYGGSRRNLGYRETY  
SEG .....  
PRD ccc

SEQ TPKPETDEMNEVETAPIPEENHVWLQPRVNRPTKPKKTSVNVYMTQVACDTKMKDRCKG  
SEG .....  
PRD ccc

SEQ STCNRYQCAGCLNWKAKIFGTLFYESSSSICRAAIHYGILDDKGLVDITRNGKVPFFV  
SEG .....  
PRD ccc

SEQ KSERHGVQSLSKYKPSSTPMVSKVKQDLDCYTTVAQLCFEKPATHCPRHCPAHCCKDE  
SEG .....  
PRD ccc

SEQ PSYWAPVFGTHIYADTSSICKTAVHAGVISNESGGVDVMPDKKTYVGSRLNGVQSES  
SEG .....  
PRD ccc

SEQ LGTPRDGKAFRIFAVRQ  
SEG .....  
PRD ccccccccccccccccc

## Prosite for DKF2phtes3\_4b4.1

PS00001	27->31	ASN_GLYCOSYLATION	PD0C00001
PS00001	41->45	ASN_GLYCOSYLATION	PD0C00001
PS00001	451->455	ASN_GLYCOSYLATION	PD0C00001
PS00004	181->185	CAMP_PHOSPHO_SITE	PD0C00004
PS00004	276->280	CAMP_PHOSPHO_SITE	PD0C00004
PS00004	464->468	CAMP_PHOSPHO_SITE	PD0C00004
PS00005	170->173	PKC_PHOSPHO_SITE	PD0C00005
PS00005	179->182	PKC_PHOSPHO_SITE	PD0C00005
PS00005	201->204	PKC_PHOSPHO_SITE	PD0C00005
PS00005	228->231	PKC_PHOSPHO_SITE	PD0C00005
PS00005	241->244	PKC_PHOSPHO_SITE	PD0C00005
PS00005	362->365	PKC_PHOSPHO_SITE	PD0C00005
PS00005	471->474	PKC_PHOSPHO_SITE	PD0C00005
PS00005	483->486	PKC_PHOSPHO_SITE	PD0C00005
PS00006	29->33	CK2_PHOSPHO_SITE	PD0C00006
PS00006	75->79	CK2_PHOSPHO_SITE	PD0C00006
PS00006	81->85	CK2_PHOSPHO_SITE	PD0C00006
PS00006	130->134	CK2_PHOSPHO_SITE	PD0C00006
PS00006	453->457	CK2_PHOSPHO_SITE	PD0C00006
PS00006	483->487	CK2_PHOSPHO_SITE	PD0C00006
PS00007	385->393	TYR_PHOSPHO_SITE	PD0C00007
PS00008	111->117	MYRISTYL	PD0C00008
PS00008	115->121	MYRISTYL	PD0C00008
PS00008	174->180	MYRISTYL	PD0C00008
PS00008	204->210	MYRISTYL	PD0C00008

WO 01/12659

PCT/IB00/01496

PS00008 227->233 MYRISTYL PDOC00008  
 PS00008 300->306 MYRISTYL PDOC00008  
 PS00008 447->453 MYRISTYL PDOC00008  
 PS00008 470->476 MYRISTYL PDOC00008  
 PS01010 195->207 SCP\_AG5\_PRI\_SC7\_2 PDOC00772

Pfam for DKFZphtes3\_4b4.1

HMM\_NAME SCP-like extracellular Proteins  
 HMM \*PQDEQDEWLNKHNDFRQOVGRGLETRGNPGPPAsNMhPMVMNDELAt  
 P + ++E+L HN +R QV P ASNM M+W+DEL +  
 Query 52 PREDKEEILMLHNKLRGGVQ-----PQASHREYMTWDELEK 88  
 HMM IAQnWANQCI FDHHDCCNHNhNPYGGNIAMWSSsTAnnPMhWssMIQNMW  
 A WA+QCI +H ++ + S GQN+ + + +++++ +Q+MY  
 Query 89 SAAANASQCInzhGPTSLVSI---GQNLGAHWG---RYASPGFHVQSWY 132  
 HMM NEVkdYNYNWNtckGG....NNFmVCGHYTONVMRnTfrIGCGRYICYC  
 +EVKDY Y + + +C HYTO+VM+ T +IGC+ C+  
 Query 133 DEVKDYTYPPSECNPMCPERCsGPMCTHYTQIVWATTNKIGCAVNTCRK 182  
 HMM HNNWcKDPWKhkhYYVCHVCPpGNtch\*  
 + M + M+ +Y VCNV P+GN+++  
 Query 183 NTVM--GEVMENAVYFVCNYSKGNWIG 208

DKFZphes3\_4f17

group: testes derived

DKFZphes3\_4f17 encodes a novel 656 amino acid protein with weak similarity to methyl-CpG-binding proteins.

Methylation at the DNA sequence 5'-CpG is required for mammalian development. Methyl-CpG-binding proteins bind specifically to methylated DNA via a related amino acid motif and can repress transcription. The novel protein does not contain such a motif.

No informative BLAST results; No predictive prosite, pfam or SCOP motif.

The new protein can find application in studying the expression profile of testis-specific genes.

similarity to methyl-CpG-binding protein

extension of HS557771/HS278337,  
there are some differences to these sequences

Sequenced by AGOWA

Locus: /map="18"

Insert length: 2320 bp

Poly A stretch at pos. 2266, polyadenylation signal at pos. 2251

```
1 GGCAGGTTCC CGGTCGCTG CGCGGGGTCG TGAGGGAGTG CGCCGGGAGC
51 GGAGATATGG AGGGAGATGG TTCAGACCCA GAGCCTCCAG ATGCCGGGGA
101 GGACAGCAAG TCCGAGAATG GGGAGAATGC GCCCATCTAC TGCACTCGCC
151 GCAAACCGGA CATCAACTGC TTCATGATCG GGTGTGACAA CTGCAATGAG
201 TGGTTCCATG GGGACTGCAAT CGGGATCACT GAGAAGATGG CCAAGGCCAT
251 CGGGAGTGGG TACTGTGGGG AGTGCAGAGA GAAGAGCCCC AACGTAGAGA
301 TTGGCTATCG GCACAAGAAG TCACGGGAGG GGGATGGCAA TGAGCGGGAC
351 AGCAGTAGGC CGCGGGATGA GGGTGGAGGG CGCAAGAGGC CTGTCCCTGA
401 TCCAGACCTG CAGCGCCGGG CAGGGTCAGG GACAGGGGTT GGGGCCATGC
451 TTGCTCGGGG CTCTGCTTCG CCCCAAAAT CTCTCCGCA GCCCTTGGTG
501 GCCACACCCA GCCAGCATCA CCAGCAGCAG CAGCAGCAGA TCAAAAGGTC
551 AGCCCCGATG TGTGTGAGT GTGAGGCATG TCGGCCCACT GAGGACTGTG
601 GTCACTGTGA TTTCTGTGG GACATGAAGA AGTTCGGGGG CCCCACAAAG
651 ATCCCGCAGA AGTGCCCGCT GCGCCAGTGC CAGCTCGGGG CCGGGGAATC
701 GTACAAGTAC TTCCCTTCCT CGCTCTCACC AGTACAGCCC TCAGACTCCC
751 TGCCAGGGCC CGCGCGGCA CTGCCACCC AACGAGGCC ACAGCATCA
801 CAGAAGTTAG GCGCATCCG TGAAGTAGG GGGCAGTGG CGTCATCAAC
851 AGTCAAGGAG CCTCCTGAGG CTACAGCCAC ACCTGAGCCA CTCTCAGATG
901 AGGACCTACC TCTGGATCCT GACTGTATC AGGACTTCTG TGCAGGGGCC
951 TTTGATACC ATGGCTGCCC CTGGATAGC GACACAGAG AGTCCCATTT
1001 CTTGAGCCCC GCGCTGCGGA AGAGGGCAGT GAAAGTGAAG CATGTGAAGC
1051 GTCGGAGAA GAAGCTGAG AAGAAGAGG AGGAGCGATA CAAGCGGCAT
1101 CGGCAGAGGC AGAAGCACAA GGATAAATGG AAACACCCAG AGAGGGCTGA
1151 TGCCAAGGAC CTTGCGTCAC TGCCCCAGTG CTTGGGGGCC GGCTGTGTGC
1201 GCGCGGCCCA GCCAGCTCC AGTATATCT CAGTAGACTG TGCCATGAAG
1251 CTGGCAGCCA ACCGCATCTA CGAGATCCTC CCCCAGCGCA TCCAGCAGTG
1301 GCAGCAGAGC CTTTGCATTG CTGAAGAGCA CGGCAGAAAG CTGCTGAAC
1351 GCATTGCGCG AGAGCAGCAG AGTGCCCGCA CCGGCTTCA GGAATGGAA
1401 CGCGGATTCC ATGAGCTTGA GGGCATCAT CTACGTGCCA AGCAGCAGGC
1451 TGTGGCGGAG GATGAGAGA GCAACAGGG TGACAGTGT GACACAGACC
1501 TGCAGATCTT CTGTGTTTCC TGTGGGCACC CCATCAACCC ACGTGTTGCC
1551 TTGGGCCACA TGGAGCGGTG CTACGCCAAG TATGAGGCC AGAGTCCTT
1601 TGGGTCCATG TACCCACAC GCATTGAAG GGGCACAGCA CTCTTCTGTG
1651 ATGTGTATAA TCCTCAGAGC AAAACATACT GTAAGCGGCT CCAGTGTGCTG
1701 TGCCCGGAGC ACTCAGCGGA CCCCAGATG CAGCTGAGC AGGTATGGG
1751 GTGCCGCCCT GTACCTGATG TCTTTGAGCT CACGGTGAC TTCTGCCGCC
1801 TGCCCAAGCG CCAGTGCAAT CGCCATTACT GCTGGAGAA GCTCGGCGT
1851 GCGGAAGTGG ACTTGGAGCG CGTGCGTGTG TGGTACAGC TGGACGAGCT
1901 GTTTGAGCAG GAGCGCAATG TCGGCACAGC CATGACAAAC CGCGCGGAT
1951 TGTGGGCGCT GATGCTGAC CAGAGCATCC AGCAGATCC CCTCATACC
2001 GACTGTGGCT CAGTGCAGCA CGCTGAGCC TCCTGGCCCG GACCCCTTAC
2051 ACCCTGCATT CCAGATGGGG GAGCGCCCG GTGGCGGTGT GTCCGTTCTT
2101 CCACTCATCT GTTCTCGGG TTCTCCCTGT GCCCATCCAC CGGTTGACCG
2151 CCCATCTGCC TTTATCAGAG GGAAGTGTCC CGTCGACATG TTCAGTGCTT
2201 GGTGGGGCTG CGGAGTCCAC TCATCTTCC CTCTCTTCCC TGGGTTTGT
2251 TAATAAAATT TTGAAGAAAC CAAAAAATA AAAAAAATA AAAAAAATA
2301 AAAAAAATA AAAAAAATA
```

BLAST Results

-----  
Entry H5557771 from database EMBL:ST:  
Human chromosome 18 clone 2 mRNA sequence.  
Score = 7582, P = 0.0e+00, identities = 1560/1598  
Entry H5278337 from database EMBL:ST:  
H.sapiens mRNA, expressed sequence tag ICRFp507H02194 (5')  
Score = 6339, P = 9.0e-281, identities = 1307/1347  
Entry H5095149 from database EMBL:ST:  
human STS W1-6941.  
Score = 1210, P = 2.2e-49, identities = 246/251

## Medline entries

98449942:  
Identification and characterization of a family of mammalian methyl-CpG  
binding proteins.  
9824997:  
Gene silencing by methyl-CpG-binding proteins.

## Peptide information for frame 3

-----  
ORF from 57 bp to 204 bp; peptide length: 656  
Category: similarity to known protein

1 MEGDGSQPEP PDAGEDSKSE NGENAPIYCI CRKPDINCFM IGCDCNENWF  
51 HGDICIRITEK MAKAIREWYC RECREKDPKL EIRYRHKKSR ERDGNERS  
101 EPRDEGGGRK RPVPDPDLQR RAUSGTGVGA MLARGSASPH KSSPQPLVAT  
151 PSQHHQOQOO QIKRSARMCG ECEACRTED CGHCDPCERDM KKGQGFNKR  
201 QKRLRQCQL RARESYKYFP SLSLPVTFPE SLPRPRPLFP TQOQOPQSK  
251 LGRIREDEGA VASSTVKEPP EATATPEPLS OEDLFLDPDL YODFCAGAFD  
301 DHGLPMSOT EESFPLDPAL RKRAVKVKHV KRREKKSEK KEERYKRHRQ  
351 KQKHKKKWKH PERADAKDPA SLPCQLGPGC VRPAQPSKY CSDDCGMKLA  
401 AMRIYEILPQ RIQWQOQSPC IAEHGGKLL ERIRREGQSA RTLQGENEAR  
451 FHELEAITLR AKQAVAEDE ESHGSGSDOT DLQITCVSGG RPIWPRVALR  
501 HNERCYAKYE SOTSFGSMYP TRIEGATRLF CDVYNPQSKT YCKRLQVLCF  
551 EHSRDPKVPV DEVCGCPVLR DVFELTGDFC RLPKRCNRH YCWEKLRRAE  
601 VDLERVRVWY KLDELFEQER NVRTAMTNRA GLALMLHQT IQHDPLTTDL  
651 RSSADR

## BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZp341f17, frame 3

TREMBL:CEFS2B11.4 gene: "F52B11.1"; Caenorhabditis elegans cosmid  
F52B11, N = 2, Score = 316, P = 8.8e-27

TREMBL:HSAB2331.1 gene: "KIAA0333"; Human mRNA for KIAA0333 gene,  
partial cds., N = 2, Score = 163, P = 2.8e-13

TREMBL:SPCC594.5 gene: "SPCC594.05c"; product: "putative  
transcriptional regulatory protein, phd finger containing"; S.pombe  
chromosome III cosmid c594., N = 3, Score = 168, P = 3.6e-12

TREMBL:AF072240.1 gene: "Mbd1"; product: "methyl-CpG binding protein  
MBD1"; Mus musculus methyl-CpG binding protein MBD1 (Mbd1) mRNA,  
complete cds., N = 2, Score = 189, P = 7.6e-11

>TREMBL:CEFS2B11.4 gene: "F52B11.1"; Caenorhabditis elegans cosmid F52B11  
Length = 523

## HSPs:

Score = 316 (47.4 bits), Expect = 8.8e-27, Sum P(2) = 8.8e-27  
Identities = 100/336 (29%), Positives = 167/336 (49%)

```
Query: 333 REKKSEKKKEERYKRRHQ-KQKHDKWKHPERADAKDASLP-OCLGPGCVRPAPQSSKY 390
      +++K+ E Y +R +Q+ D + +A +P P QCL P C+ ++ SKY
Sbjct: 118 QQRKANIINERDYPNRPPTQGSADLRARQTQNA-EPDKHPRQCLNPNCIYESRDSKY 176

Query: 391 CSDDCGMKLAANRIYEILPQRIQQW-----QSPCTAEHGGKLLERIRREQSARTLQ 445
      CSD+CG +LA R+ EILP R +Q+ P E+ K +I RE Q +
Sbjct: 177 CSDECGKELANRLTEILPNRCKQYFFEGPSGGPSLEDEIKPKRAKINREVQKLTSEK 236

Query: 446 EMERRFHEL-EAIIILRAKQAVREDEESNEGDSDDTLQIFCVSCGHPINPRVAL-RHME 503
      M ++L E I + K Q + +E D +L C+ CG P P + +H+E
Sbjct: 237 NMMAFLNKLVEFIKTLQKLQPLGTEERY-----DDNLYEGCIVGLPDIPLKRYTKHIE 290

Query: 504 RCYAKYESQTSFGSMYPTRIEGATRLFCQVYNPQSKTYKRLQVLCPEHSRDPKVPADV 563
      C+A+ E SFG+ P + +C+ Y+ ++ ++CKRL+ LCPEH + +V
Sbjct: 291 LCMARSEKAIISFGA--PEK--NNDMFYCEKYDSRTNSFCRLKSLCPEHRKLGDEQHLKV 346

Query: 564 CGCP-----LVRDVFELTGF----CRLPKRCNRHYCEKLRRAEVDLVRV 607
      CG P V ++ E+ F CR K C++H+ W R ++LE+
Sbjct: 347 CGYPKKWEDQMISTAKTVSELIEMEDPFGEGRGRTKKDACHKHKWIPSLRGITIELEQAC 406

Query: 608 VMYKDELFEQ--ERNVRTANTNRAGLLALMLHOTIQHNDPLTTDLRSSA 654
      ++ K+ EL + + H T A L+M+H+ + + LR+ A
Sbjct: 407 LFKMYELCHEMHLNAHAEMTTNA--LSIMHMKQPSTKCSFFLRNFA 453

Score = 53 (8.0 bits), Expect = 8.8e-27, Sum P(2) = 8.8e-27
Identities = 24/100 (24%), Positives = 41/100 (41%)

Query: 169 CGCEACRRTEDCGHCDPCR-----DMKK-FGGPNKIROKRLRQCOLRARESYKYPSS 222
      C C C ++CG C CR DM+K F +K + RQ + + +
Sbjct: 17 CMNCIRCNDKNCGTWPCRNKGTCDMRKCFSAKRLYNEKV-ROTDENLK-AIMAKTAQ 74

Query: 223 LSPVTPSESLPRPRPLPTQOQPOPSQKLGRIR-EDEGAVASS 264
      + + P + +QO + +K GR + G A+
Sbjct: 75 REAAHQAAATTAPSAPVIEQOVE-KKRGKRGSGNGGAA 116

Score = 48 (7.2 bits), Expect = 2.9e-26, Sum P(2) = 2.9e-26
Identities = 13/39 (33%), Positives = 19/39 (48%)

Query: 179 EDCGHCDPCRDMKKFGG--PNKIROKRLRQCOLRARESY 216
      E C +C C D K G P + + C +R+ C A+ Y
Sbjct: 15 ERCHNCIRCNDKNCGTWPCRNKGTCDMRKCFSAKRLY 53
```

Pedant information for DKFZphtes3\_4f17, frame 3

Report for DKFZphtes3\_4f17.3

```
[LENGTH] 656
[MM] 75711.71
[pI] 8.61
[HMOL] TREMBL:CEF52B11_4 gene: "F52B11.1": Caenorhabditis elegans cosmid F52B11 3e-25

[FUNCTION] 99 unclassified proteins [S. cerevisiae, YPL138c] 3e-10
[FUNCTION] 04.05.01.04 transcriptional control [S. cerevisiae, YNL037c] 2e-04
[PROSITE] MYRISTYL 6
[PROSITE] AMIDATION 2
[PROSITE] CK2_PHOSPHO_SITE 8
[PROSITE] TYR_PHOSPHO_SITE 3
[PROSITE] GLYCOSAMINOGLYCAN 1
[PROSITE] PKC_PHOSPHO_SITE 9
[KW] All_Alpha
[KW] LOW_COMPLEXITY 18.75 %
[KW] COILED_COIL 4.57 %

SEQ MEGDGSDEPPDAGEDSKSENGENAPIYICIRKPDINCFMIGCDNCFNHGDCIRITEK
SEG .....
PRD ccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc
COILS .....

SEQ NAKAIREMYCREKOPKLEIRYRHKRSERDGNRDSSEPRDEGGGRKRPVDPOLQR
SEG .....
PRD hhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh
COILS .....

SEQ RAGSGTGVAHLAGSASPHKSSPOPLVATPSQHQQOQQQIKRSARNCGEACARTED
SEG .....
PRD cccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc
COILS .....
```

[illegible]

Prosite for DKFZphtes3\_4f17.3

PS000002	124-1128	GLYCOSAMINOGLYCAN	PC00000002
PS000003	58-961	PKC_PHOSPHO_SITE	PC00000003
PS000005	165-618	PKC_PHOSPHO_SITE	PC00000005
PS000005	215-218	PKC_PHOSPHO_SITE	PC00000005
PS000005	245-248	PKC_PHOSPHO_SITE	PC00000005
PS000005	268-269	PKC_PHOSPHO_SITE	PC00000005
PS000005	337-340	PKC_PHOSPHO_SITE	PC00000005
PS000005	387-390	PKC_PHOSPHO_SITE	PC00000005
PS000005	439-442	PKC_PHOSPHO_SITE	PC00000005
PS000005	627-630	PKC_PHOSPHO_SITE	PC00000005
PS000006	6-10	CK2_PHOSPHO_SITE	PC00000006
PS000006	15-21	CK2_PHOSPHO_SITE	PC00000006
PS000006	227-231	CK2_PHOSPHO_SITE	PC00000006
PS000006	265-269	CK2_PHOSPHO_SITE	PC00000006
PS000006	280-284	CK2_PHOSPHO_SITE	PC00000006
PS000006	308-312	CK2_PHOSPHO_SITE	PC00000006
PS000006	521-525	CK2_PHOSPHO_SITE	PC00000006
PS000006	652-656	CK2_PHOSPHO_SITE	PC00000006
PS000007	339-346	TYR_PHOSPHO_SITE	PC00000007
PS000007	1007-1010	TYR_PHOSPHO_SITE	PC00000007
PS000007	211-219	TYR_PHOSPHO_SITE	PC00000007
PS000008	42-48	MYRISTYL	PC00000008
PS000008	123-129	MYRISTYL	PC00000008
PS000008	131-131	MYRISTYL	PC00000008
PS000008	129-135	MYRISTYL	PC00000008
PS000008	259-265	MYRISTYL	PC00000008
PS000008	396-402	MYRISTYL	PC00000008
PS000009	107-111	AMIDATION	PC00000009
PS000009	425-429	AMIDATION	PC00000009

(No Pfam data available for DKFZphtes3\_4f17.3)

DKFZphtes3\_4f5  
-----

group: signal transduction

DKFZphtes3\_4f5.3 encodes a novel 790 amino acid protein similar to beta-transducins.

The protein contains 3 WD-40 repeats, which are typical for the beta-transducin subunit of G-proteins. The beta subunits seem to be required for the replacement of GDP by GTP as well as for membrane anchoring and receptor recognition. In addition, a Cytochrome C family heme-binding site signature is present. The protein is larger (790 amino acids) than the usual eukaryotic G-beta transducins (about 340 amino acids).

The new protein can find application in modulating/blocking G-protein-dependent pathways.

similarity to S.pombe "beta-transducin"

complete cDNA, EST hits  
complete cds,  
on genomic level encoded by HS313D11, at least 7 exons these exons  
match  
only partially with the predicted transcripts in HS313D11

Sequenced by AGOWA

Locus: /map="16p13.3"

Insert length: 3166 bp

No poly A stretch found, no polyadenylation signal found

```
1 GCGGGCTTCC GCGCGGGGGG TTCCGGACAA CCGTGGCGTT TTAGTAAAG
51 ATTGGGGTTC GCGCGGGGGG GAAGGGCTGC CCGGGGCCCT CTGGTCTCTG
101 TCCCGGAGGG TCCGCTCCCC CGCGCCACTG CGCGGCTCCC AGGAACCTTG
151 TACTCCGGGG TCGGCGGCTT CTCTCTGACC TCGGGTCCGG CAGACACACT
201 CGAGCTCTTT AAGTAGCTCG GTCCCTTGACG TCCCTCTGGG CCCTTCCCGC
251 GTCTATCGCC TGAGTCCCCG GCGCCCTCTA GCGCTCTGTT CCTCCCTCTC
301 TTTTGTCTCT CCTAGAGCC CCGCGGCCCT CAGGGCTGAC AGTGTGGAGG
351 GCGGGAGTCT CTTCCGCTCC CTGCTGGCAT TGACTGACCG AGCGTTTAGT
401 GACTGGCCAG ATCTGGCTGA TGGGGGTACC GAGAGTGGC CTGGGCCCGG
451 AATGTCCAGC TAGAGTCTTC CGTGGAAAGT AGACATGAAG CTGACAGGCC
501 TAAGGGAAGC TAGGAAGTCC CTTACCGCTC CAGCCAGGGT GATGGGCTGG
551 ACTGACAGAC TCCAGTGAAT TTGAGCTTGC CTGTCAAGCT GATTGGCTGA
601 TAGACAGCCC TGGATTGGCT CACTAAGACT GACCAAGCCG GGACCAAGCA
651 GTCTGGGGGT CCAACCTGGG GTGGAAAGTC TGAACCTGAT ACCCAACCCG
701 GCTGACCAAG CCAGCCCAAC TGACTGACCT CCGTACCCCT GACCTCATCA
751 CCTGTGACAG CATGGAGAAG ATGTCCCGTG TGACCACAGC CCGGGTGGGC
801 AGCGTGTCTG CAGGCGCGAC CATGCACTGC CACTGGATG CTCGCGCCAA
851 TGCCATGAGT GTGTGCGCGG ACGCAGCCCA GTGTGCTGT GCGAGGCGTA
901 GCATCTTCAA GATCTATGCC ATCGAGGAGG AACAGTTGCT GGAAGAGCTG
951 AACCTGCGTG TGGGGCGCAA GCCTTCGCTT AACCTGACCT GTGTGACGTT
1001 GGTCTGGCAC CAGATGGATG AGAACCTGCT GGCCACAGCA GCCACCAATG
1051 GCGTGGTGGT CAGGTGGAAC CTGGGCGCGC CATCCCGCAA CAGCAGAGAC
1101 CAGCTGTTCG CAGACACAAA GCGCACGGTA AACAAAGTCT GCTTCCACCC
1151 CACCGAAGCC CAGCTGCTGC TCAGTGGGTC CCAGGATGGC TTCATGAAGT
1201 GCTTTGACCT CCGCAGAAAG GACTCTGTCA GCACCTTCTC GGGCCAGTGG
1251 GAGAGCGTGC GGGACGTGCA GTTCAGTATC CCGGACTACT TCACCTTCGC
1301 CTCCAGCTTT GAGAACGGCA ATGTGCACTC CTGGGACATC CGCGTCCCG
1351 ACCGCTGGCA GAGGATGTTT ACAGCCCAAC ACAGCCCGCT CTTCGTCTGC
1401 GACTGGCACC CCGAGGACAG GGGCTGGTTG GCGACTGGAG GCGCGACAAA
1451 GATGGTGAAG GTCTGGGACA TGACCACGCA CCGTGCCAAAG GAGATGCACT
1501 GTGTGCAAGC CATCGGCTCG GTGGCCGCTG TGAAGTGGCG GCCAGAGTGC
1551 CGGCACACCC TGCCACGCTG CTGCATGATG GTGGACACCA ACATCTATGT
1601 TTGGGACGTG CGCCGGCCCT TCGTGGCAGC TGCCATGTTT GAGGAACACC
1651 GAGAGCTCAC CACGGGAATT GCGTGGCGCC ACCCCACAGA CCGCTCTCTC
1701 CTGCTGTCTG GCTCCAAAGG CAGCTCGCTG TGCCAGCACC TGTTCGCGGA
1751 CGCAGGCCAG CCGCTCGAGC GCGCCAAACC TGAGGCGCTC TGCTACGGCC
1801 TCTTGGGGGA CCGTGGCTTC GCGCCAAAGG AGAGCTGCTG GCGTCCGCGG
1851 TCGGGGCGCA AGCCCTACAC TGGCGACCGG CGCCACCCCA TCTTCTTTAA
1901 GCGCAAGCTG GACCTGCGCG AGCCCTTCGC AGGCTCTGCC TCCAGTGGCC
1951 TCAGTGTCTT TGAGACGGAG CAGGTGGCGG GCGGCATGGG CTGGTGTGTTG
2001 GACACAGCTG AGCGTTATGC GCTGGCTGGC CGGCCACTGG CCGAGCTCTG
2051 TGACACCAAC GCAAGAGTGG CTCGAGAGCT TGCGCGCAAC CAGGTGGCGG
2101 AAACGTGGAC CATGCTGGCG ATCATCTACT GCAGCCCTGG CCTAGTGGCC
2151 ACTGCAAAAC TCAACACAGC TGTGGGCAAG GTGGCTCTCT GTGGCTCTCCC
2201 GCTCATGAAC AGTTTCAACC TGAAGGATAT GCGCCCAAGG TTGGGCAATG
2251 AGACCGGCTG GACCGCAGC AAAGGAGATG CAGCGAGGCA CACAGTTCTG
2301 CTGCTCTCTC CGGCCACACT CATACCAAT GAGGATAACG AGGAACCGGA
2351 GGGCAGCGAC GTACCTGCGC ACTACCTGAT GGGTACAGTG GAAGTGAGG
```



```
2401 AGGACGAGCT GTACCTGCTG GATCCGGAAC ACCGGCACCC CGAGGACCGT
2451 GAGTCCCTCC TGCCCGCAGG GGCCTTTCCG CTGGCCACAG AGATCGTGGA
2501 CACGCCCTCC GGACCCGAGC ACCTGCAGGA CAAGGCCGAC TCCTCCGACG
2551 TGAGCCGCGC CGAGCGGAT GTGGCTCCG TGGCCCGCT GGACTCTCC
2601 TTCTGGCTCC TGTCTGCTC ACACGGCTC TACGACGCC GCCTGCCGCC
2651 CGACTTCTTC GCGTGTGG TGGCGGCAT GCTGCCTTC TACGTGAGC
2701 AGGGCGACGT GCAGATGGCT GTGTCTGTC TCATGCTCTT GGTGAACGG
2751 GTGGCAAGG ACATCGACGA GCAGACCCAG GAGCACTGCT ACATCTCTTA
2801 CATCGACCTG CTGACGCTT TCCGCTCTG GAGCTGTCC ACGAGGTGG
2851 TCAGCTGAG CACGAGCCGC GCGCTGAGT GCCTCAACCA GGCCTCCACC
2901 ACCCTGCAGG TCAACTGCAG CCACTGCAG GCGCCATGA GCAGCCGGGG
2951 CTGGGTCTGC GACAGTGCC ACCGTGCGC CAGATGTGT GCCGTCTGCC
3001 ACCACGTAGT CAAGGGTCTC TTCTGTGCT GCAAGGGCTG CAGCCACGG
3051 GGCACCTGC ASCACATCAT GAGTGGCTG GAAGCGAGT CCACTGTCC
3101 CGAGCGCTGC GGCACCTCT GCGAGTACT CTGACGGGG ATCTGCTGGG
3151 CTTGCCCGGG CGGCCG
```

## BLAST Results

Entry HS313D11 from database EMBL:  
Human DNA sequence from cosmid 313D11 from a contig on the short arm of  
chromosome 16. Contains ESTs, STS and CpG islands.  
Score = 6238, P = 0.0e+00, identities = 1318/1391

## Medline entries

No Medline entry

## Peptide information for frame 3

ORF from 762 bp to 3131 bp; peptide length: 790  
Category: similarity to known protein

```
1 MEKMSRVTTA LGGSVLTGRT MHCHLDAPAN AISVCRDAAQ VVVAGRSIFK
51 IYIEEEQFV EKLNLVGRK PSNLSCADV VWRMDENLL ATAATNGVVV
101 TWNLGRPSRM KQDLPTENK RTVHKVCFHP TEARVLSSG QDGTNKEFDL
151 RKQSVSTVS QGSESVROVQ PSIRDYITPA SYFENGVOQL WDIRRPRDCE
201 RMFTAHNGPV FCCDWHFEDR GMLATGGRDK MVKWMQMTI RAKENHCVQT
251 IASVARVKWR PEGRHHLATC SHNVHNIYV WDVRRPFVPA AMFEHRDVT
301 TGIAMRHPRD PSFLLSGSKD SSLCOHLFRD ASQVPERANP EGLCYGLFUD
351 LAFAAKESLV AAESGRKPYT GORNRPIFK RLIDPAEPA GLASSALSVT
401 ETEPGGGGHR WFYDTAERYA LAGRPLAELC DHNAKVAREL GRNOQAOTWT
451 MLRIIYCSPG LVPTANLHNS VCKGGSCGLP LMSFNLKDM APGLGSETRL
501 DRSGDARSR TVLLDSSATL ITNEDNEETE GSDVPADYLL GDVEGEDEEL
551 YLLOPEHARP EDPECVLPQE AFLRHEIVD TPQPEHLAD KADSPHYVSG
601 EADVASLAPV DSSLSLVSU HALYDSRLPP DFTVLVLRDM LHFYASGQDV
651 QMAVSVLIVL GERVRKIDE QTOEHMYTSY IDLQRFRLM NVSNEVVKLS
701 TSRAVSLCMLQ ASTTLLVNCS KCKRPMSSSY WVCORCHRA SMCVACHHV
751 KGLFVVCQCG SHGGHLQHIM KWLEGSSHCP AGCCHLCEYS
```

## BLASTP hits

Entry YD58\_SCHPO from database SWISSPROT:  
HYPOTHETICAL 93.2 KD TRP-ASP REPEATS CONTAINING PROTEIN C4F8.11 IN  
CHROMOSOME 1. >TREMBL:SPAC4F8.11 gene: "SPAC4F8.11"; product:  
"beta-transducin"; S.pombe chromosome 1 cosmid c4f8.  
Score = 404, P = 3.0e-42, identities = 169/639, positives = 278/639

Entry PEX7\_HUMAN from database SWISSPROT:  
PEROXISOMAL TARGETING SIGNAL 2 RECEPTOR (PTS2 RECEPTOR) (PEROXIN-7).  
>TREMBL:HSU76560.1 gene: "Pex7"; product: "peroxisome targeting signal  
2 receptor"; Human peroxisome targeting signal 2 receptor (Pex7) mRNA,  
complete cds. >TREMBL:HSU88871.1 gene: "HsPEX7"; product: "HsPex7p";  
Human HsPex7p (HsPEX7) mRNA, complete cds.  
Score = 220, P = 1.1e-15, identities = 62/244, positives = 107/244

Entry PEX7\_MOUSE from database SWISSPROT:  
PEROXISOMAL TARGETING SIGNAL 2 RECEPTOR (PTS2 RECEPTOR) (PEROXIN-7).  
>TREMBL:M46917.1 product: "peroxisomal PTS2 receptor"; Mus musculus  
peroxisomal PTS2 receptor mRNA, complete cds.  
Score = 214, P = 5.3e-15, identities = 60/240, positives = 106/240

Entry ATAC2294\_7 from database TREMBL:  
 gene: "FliP17.7": Arabidopsis thaliana chromosome I BAC FliP17 genomic  
 sequence, complete sequence.  
 Score = 232, P = 3.4e-14, identities = 68/260, positives = 120/260

Entry S66835 from database PIR:  
 probable membrane protein YOL138c - yeast (Saccharomyces cerevisiae)  
 >TREMBL:SCYOL138C\_1 S.cerevisiae chromosome XV reading frame ORF  
 YOL138c  
 Score = 136, P = 2.5e-13, identities = 24/77, positives = 44/77

Alert BLASTP hits for DKFZphtes3\_4f5, frame 3

No Alert BLASTP hits found

Pedant information for DKFZphtes3\_4f5, frame 3

Report for DKFZphtes3\_4f5.3

[LENGTH] 790  
 [MM] 88207.10  
 [PI] 5.05  
 [HOMOL] SWISSPROT:YDSB\_SCHPO HYPOTHETICAL 93.2 KD TRP-ASP REPEATS CONTAINING PROTEIN  
 C4F8.11 IN CHROMOSOME I. 9e-44  
 [FUNCAT] 99 unclassified proteins [S. cerevisiae, YOL138c] 5e-16  
 [FUNCAT] 10.04.09 regulation of g-protein activity [S. cerevisiae, YBR195c] 3e-11  
 [FUNCAT] 06.10 assembly of protein complexes [S. cerevisiae, YBR195c] 3e-11  
 [FUNCAT] 03.16 dna synthesis and replication [S. cerevisiae, YBR195c] 3e-11  
 [FUNCAT] 09.13 biogenesis of chromosome structure [S. cerevisiae, YBR195c] 3e-11  
 [FUNCAT] 04.05.01.07 chromatin modification [S. cerevisiae, YBR195c] 3e-11  
 [FUNCAT] 30.10 nuclear organization [S. cerevisiae, YCR072c beta-transducin family]  
 3e-10  
 [FUNCAT] 04.05.01.01 general transcription activities [S. cerevisiae, YBR198c  
 TAF90 - TFIID subunit] 9e-09  
 [FUNCAT] 04.01.04 rna processing [S. cerevisiae, YLL011w] 1e-07  
 [FUNCAT] 30.09 organization of intracellular transport vesicles [S. cerevisiae,  
 YDL195w] 2e-07  
 [FUNCAT] 08.07 vesicular transport (golgi network, etc.) [S. cerevisiae, YDL195w]  
 2e-07  
 [FUNCAT] 30.19 peroxisomal organization [S. cerevisiae, YDR142c] 4e-07  
 [FUNCAT] 06.04 protein targeting, sorting and translocation [S. cerevisiae, YDR142c]  
 4e-07  
 [FUNCAT] 08.10 peroxisomal transport [S. cerevisiae, YDR142c] 4e-07  
 [FUNCAT] 08.01 nuclear transport [S. cerevisiae, YER107c] 4e-07  
 [FUNCAT] 04.07 rna transport [S. cerevisiae, YER107c] 4e-07  
 [FUNCAT] 30.03 organization of cytoplasm [S. cerevisiae, YER107c] 4e-07  
 [FUNCAT] 03.22 cell cycle control and mitosis [S. cerevisiae, YGL003c] 5e-07  
 [FUNCAT] 06.13 proteolysis [S. cerevisiae, YGL003c] 5e-07  
 [FUNCAT] 04.05.01.04 transcriptional control [S. cerevisiae, YCR084c] 8e-07  
 [FUNCAT] 04.05.03 rna processing (splicing) [S. cerevisiae, YPR178w] 1e-06  
 [FUNCAT] 03.13 meiosis [S. cerevisiae, YLR129w] 3e-06  
 [FUNCAT] 03.25 cytokinesis [S. cerevisiae, YCR057c] 1e-05  
 [FUNCAT] 03.04 budding, cell polarity and filament formation [S. cerevisiae, YCR057c]  
 1e-05  
 [FUNCAT] 06.07 protein modification (glycosylation, acylation, myristylation,  
 palmitoylation, farnesylation and processing) [S. cerevisiae, YEL056w] 2e-04  
 [FUNCAT] 30.04 organization of cytoskeleton [S. cerevisiae, YOR272w] 6e-04  
 [SCOP] dlgothb\_2.46.3.1.1 betal-subunit of the signal-transducing 5e-06  
 [PIRKN] duplication 7e-10  
 [PIRKN] signal transduction 7e-08  
 [PIRKN] peroxisome 9e-06  
 [PIRKN] heterotrimer 7e-08  
 [PIRKN] GTP binding 7e-08  
 [PIRKN] peroxisome biogenesis 9e-06  
 [PIRKN] transmembrane protein 1e-14  
 [SUPFAM] MS11 protein 7e-10  
 [SUPFAM] WD repeat homology 1e-14  
 [SUPFAM] GTP-binding regulatory protein beta chain 7e-08  
 [SUPFAM] PR11 protein 3e-08  
 [SUPFAM] coatamer complex beta' chain 1e-06  
 [PROSITE] CYTOCHROME\_C 1  
 [PROSITE] WD\_REPEATS 3  
 [PROSITE] MYRISTYL 10  
 [PROSITE] AMIDATION 2  
 [PROSITE] CAMP\_PHOSPHO\_SITE 2  
 [PROSITE] CK2\_PHOSPHO\_SITE 11

```

[PROSITE] TYR_PHOSPHO_SITE 1
[PROSITE] PKC_PHOSPHO_SITE 7
[PROSITE] ASN_GLYCOSYLATION 4
[PFAM] WD_domain, G-beta repeats
[KW] Ali_Beta
[KW] 10
[KW] LOW_COMPLEXITY 2.28 %

```

```

SEQ MEKHSRVTTALGGSVLTGRTHCHLQAPANAISVCRDAAQVVVAGRSIFKIYAIIEEQFV
SEG .....
IgotB .....
SEQ EKLNLRVGRPSNLNLSCADVVMQMDENLLATAATNGVVVTWNLGRPSRNKQDQLFTEHK
SEG .....
IgotB .....TTCEEEEEETTTTTEEEET-TTCEEE--EEECCE
SEQ RTVNKVCFHPTFAHVLGSGQDGFNCFDLARKKDSVSTFSGQSESVRDVQFSIRDYTFFA
SEG .....
IgotB .....CCEEEEEETTTTCEEEEEETTTTTEEEETTTTCEEEEEETTTTTEEE
SEQ STFENGNVQLMDIRRPORCERMTAHNGPVFCDDWHPEDRGWLATGGDKMKVVMWMTTH
SEG .....
IgotB .....E-ETTTTEEEETTTTTEEE-EEECCEEEEEETTTTCEEEEEETTTTTEEE
SEQ RAKENHCVCQTIASVARVWRPECRHHLATCSMHVDHNIYVMDVRRFPVPAAMFEHRDVT
SEG .....
IgotB .....
SEQ TGIAMRHPHDPFLLSGSKDSSLCQHLFRDASQPERANPEGLCYGLFGDLAFPAKESLV
SEG .....
IgotB .....
SEQ AAESGRKPYTGDRNHPFFKRLDPAEPFAGLASSALSVFETEPGGGHRMFVDTAERYA
SEG .....
IgotB .....
SEQ LAGRLAELCOHNAKVAELGRNQVAQTWNLRIIYCSPGLVPTANLHNSVKGSGCGLP
SEG .....
IgotB .....
SEQ LKNSFNLRDMAPGLGSETRLDKSGDARSQTVLLOSSATLITNEONEETEGSDVPADYLL
SEG .....
IgotB .....
SEQ GDVEGEDELYLLDPEHARPEDECVLPQEAFFLRAHEIVDTPGPEHLQDKADSPHVSQS
SEG .....
IgotB .....
SEQ EADVASLAPVDSSFSLLSVSHALYDSRLPDPFFGLVRDMLHFFYAEQGDVQMAVSLLIVL
SEG .....
IgotB .....
SEQ GERVRKIDIEQTEHMYTSDYDLLQRFRLMNVSNVVKLSTSRVSCNLQASTTLHVMCS
SEG .....
IgotB .....
SEQ HCKRPMSSRGWVCDRCRCASMCVCHHVKGLEFVWCGCCHGGLQHIKXWLGSSHCP
SEG .....
IgotB .....
SEQ AGCGHLCEYS
SEG .....
IgotB .....

```

## Prosites for DKFZphtes3\_4f5.3

PS00001	74->78	ASN_GLYCOSYLATION	PD0000001
PS00001	468->472	ASN_GLYCOSYLATION	PD0000001
PS00001	691->695	ASN_GLYCOSYLATION	PD0000001
PS00001	718->722	ASN_GLYCOSYLATION	PD0000001
PS00004	69->73	CAMP_PHOSPHO_SITE	PD0000004
PS00004	152->156	CAMP_PHOSPHO_SITE	PD0000004
PS00005	17->20	PKC_PHOSPHO_SITE	PD0000005
PS00005	165->168	PKC_PHOSPHO_SITE	PD0000005
PS00005	172->175	PKC_PHOSPHO_SITE	PD0000005
PS00005	239->242	PKC_PHOSPHO_SITE	PD0000005
PS00005	364->367	PKC_PHOSPHO_SITE	PD0000005
PS00005	701->704	PKC_PHOSPHO_SITE	PD0000005

WO 01/12659

PCT/IB00/01496

PS00005	727->730	PKC_PHOSPHO_SITE	PDOC00005
PS00006	76->80	CK2_PHOSPHO_SITE	PDOC00006
PS00006	165->169	CK2_PHOSPHO_SITE	PDOC00006
PS00006	172->176	CK2_PHOSPHO_SITE	PDOC00006
PS00006	181->185	CK2_PHOSPHO_SITE	PDOC00006
PS00006	398->402	CK2_PHOSPHO_SITE	PDOC00006
PS00006	498->502	CK2_PHOSPHO_SITE	PDOC00006
PS00006	503->507	CK2_PHOSPHO_SITE	PDOC00006
PS00006	522->526	CK2_PHOSPHO_SITE	PDOC00006
PS00006	598->602	CK2_PHOSPHO_SITE	PDOC00006
PS00006	600->604	CK2_PHOSPHO_SITE	PDOC00006
PS00006	675->683	CK2_PHOSPHO_SITE	PDOC00006
PS00007	337->346	TYR_PHOSPHO_SITE	PDOC00007
PS00008	13->19	MYRISTYL	PDOC00008
PS00008	97->103	MYRISTYL	PDOC00008
PS00008	139->145	MYRISTYL	PDOC00008
PS00008	161->167	MYRISTYL	PDOC00008
PS00008	317->323	MYRISTYL	PDOC00008
PS00008	342->348	MYRISTYL	PDOC00008
PS00008	391->397	MYRISTYL	PDOC00008
PS00008	460->466	MYRISTYL	PDOC00008
PS00008	474->480	MYRISTYL	PDOC00008
PS00008	759->765	MYRISTYL	PDOC00008
PS00009	67->71	AMIDATION	PDOC00009
PS00009	364->368	AMIDATION	PDOC00009
PS00190	743->749	CYTOCHROME_C	PDOC00169
PS00678	90->105	WD_REPEATS	PDOC00574
PS00678	223->238	WD_REPEATS	PDOC00574
PS00678	269->284	WD_REPEATS	PDOC00574

Pfam for DKFZphtes3\_4f5.3

HHH_NAME	WD domain, G-beta repeats		
HHH	*MKGHnnMVMCVaFSPDGnWFiVSGSNdgTCRLWD*		
	** HN+V C+ ++P+ R +++G++D+ +++WD		
Query	203	FTAHNGPVFCDDWHPEDRGWLATGGRDKHVKVWD	236

DKFZphtes3.4h6

group: intracellular transport/trafficking

DKFZphtes3.4h6 encodes a novel 622 amino acid protein with strong similarity to the kinesin light chain.

Kinesin is a microtubule-based motor protein that pulls vesicles or organelles towards the plus end of microtubules. Structural changes in the protein that drive motility are coupled to ATP binding and hydrolysis. The novel protein is similar to kinesin light chain, which is part of the functional kinesin holoenzyme tetrameric protein. The light chain has been proposed to function in coupling of cargo to the heavy chain or in the modulation of the ATPase activity of the heavy chain. The novel protein contains two kinesin light chain repeats and one RGD cell-attachment site.

The novel kinesin protein can find application in modulating the function of kinesin and modulating intracellular transport via/on microtubules.

strong similarity to Kinesin light chain

complete cDNA, complete cds, start at 150, EST hits (few)

Sequenced by AGOWA

Locus: unknown

Insert length: 2992 bp

Poly A stretch at pos. 2914, polyadenylation signal at pos. 2893

```
1 GCGGGGATGG AGCGGGGGG ACCGGCTCGG GGGTGGGGT CCGGGTGAAG
51 CGGGAGGCGG CCAGAGTCGG AGCGGGGGCC GAGCACCAGG CGCAGGCGCG
101 GCGCCCGGCT GCGCGCACCC TCGTCTCTAC AGACGCCACA GCCATGGCCA
151 TGATGGTATT TCCGCGGGAG GAGAAGCTCA GCCAGATGA GATCGTGTG
201 GGCACCAAGG CTGTCTATCA GGGACTGGAG ACTGTGCTG GGAAGCATG
251 TGCCCTGCTG GCTCTCTGCG TTGACCTGGA GCGCGGCGAA CGCAGCCCTG
301 GCTCGGAGGA GCGCTGCATC CTCTGCGTC GCTCCCTGGA AGCATTGAG
351 CTGGGGCTGG GGGAGGCCCA GGTGATCTTG GCATTGTGGA GCCACCTGGG
401 GGCTGTAGAA TCAGAGAAAC AGAAGCTGGG GCGCAGGTG CGCGTCTGG
451 TCGAGGAGAA CCACTGCTGT CTGAGGAGC TGGCGGGGAC ACACAGAG
501 CTGACGCGCA GTGAGCAGGC CGTGGCCGAC CTGAGGAGG AGAAGCAGCA
551 CTGTCTGTTT ATGAGCCAGA TCGCAAGTT GGATGAAGAC GCCTCCCTTA
601 ACGAGGAGAA GGGGAGCTGC CCCAAGACA CACTGGATGA CCGTCTCCCG
651 AATGAGGATG AGCAGAGCCC AGCCCTTAGC CCAGAGGAG GGAATGTTC
701 TGCTCAGCAT GGGGGCTAGC AGATCCCGGC CGCGCTCCGC ACCTGTGACA
751 ACCTGGTGAT CCAATACGCC TCACAGGGCC GCTACAGGT AGCTGTGCCA
801 CTCTGCAAGG AGGCACTCGA AGACCTGGAG AAGACGTGAG GCCACGACCA
851 CCCTGACGTT GCCACCATGC TGAACATCCT GGCACGTGTC TATCGGGATC
901 AGAACAGATA CAAGGAGGCT GCGCACCTGC TCAATGATGC TGTGGCATC
951 CGGGAGAAAA CACTGGGCAA GGACACCCCA GCGGTGGCTG CGACACTAAA
1001 CAACCTGGCA GTCTGTATG GCAAGAGGGG CAAGTACAAG GAGGCTGAGC
1051 CATTGTGCAA GCGGGCACTG GAGATCCGGG AGAAGGTCTT GGGCAAGTTT
1101 CACCCAGATG TGGCCAGACA GCTGAGCAAC CTGGCCCTCG TTGCGCAGAA
1151 CCAGGGCAAA GGTGAGGAGG TGGATATTA CTATCCCGCG CCACTGGAGA
1201 TCTATGCTAC ACGCTCGGG CCGATGACC CCAATGTGGC CAAGACCAAG
1251 AACACCTGGG CTCTCTGCTA CCTGAAGCAG GCAAGTACCC AGGATGCGGA
1301 GACCTTGTAC AAGGAGATCC TCACCCGCGC TCATGAGAAA GAGTTTGGCT
1351 CTGTCAATGG GGACACAGAG CCCATCTGGA TGACGAGA GGAGCGGAG
1401 GAAAGCAAGG ATAGCGCCCG GGACAGCCCG CCCTATGGGG AATACGGCAG
1451 CTGGTACAAG GCCTGTAAAG TAGACAGCCC CACAGTCAAC ACCACCTGCG
1501 GCAGCTTGGG GCGCCTATAC CGCGCCGAGG GCAAGCTGGA AGCCGCGCAC
1551 ACACCTAGAG ACTGTGCCAG CCGTAACCCG AAGCAGGTTT TGGACCCCGC
1601 AAGCCAGACC AAGTGTGAG AACTCTGAAA AGATGGCAGT GGCAGGGGGG
1651 GAGACCCCGC CAGCAGCCGA GACATGGCTG GGGGTGCCGG GCCTCGGTCT
1701 GAGTCTGACC TCGAGGAGCT GGAAGCTACA GCTGAGTGA ATGGGGATGG
1751 CAGTGGCTCC TTGAGGCGCA CGGTTCTCTT TGGGAACTC CGGGATGCCC
1801 TGAGGCGCAG CAGTGAAGTG CTGTAAAGA AGCTGCAAGG GGGCACCCCC
1851 CAGGAGCCCG CTAACCCGAG GATGAAGCGG GCAAGTTGCG TCAACTCTCT
1901 CAACAAGAGC GTGGAAGAGC CGACCCAGCC TGAAGCCACA GGTCTCTCTG
1951 ACAGCCCGAC TCTCAGCTCC AGCTCCATGG ACCTCTCCCG ACGAAGCTCC
2001 CTGGTGGGCT AATGCTGAAG GGGCAGCCAG TCACAGAGG GCCACCTGG
2051 CACACCCCCC TCACCCGAGC CCGTGCATG GGCCTGTGTC TTGTCCCGCC
2101 TGTCTCTCCG ACAGCCCTG TCTTTCTGT TCAATCTCAG GGTAACTTCT
2151 TCCCTTGTCA TCTCAGCTG AGCCCTGGAG GCTGGCCCTG CCCACTCCAG
2201 CTCATCTCCCT TATTATTCTC TTCCAGCAGG GCGCTCTTCC CTAGGTTCCG
2251 GCCAGCAGGA GGTGCGGCTG GGAGTCTCCA CCAATAGACT AGTGGCCTGG
2301 CTTCCGCGAG CCCCAGAGC AAGAACACTA AGCACTCCCG GCGCTCTGG
2351 CACCTTCGCC CTCCCTCCG ACTCAACCCG GCGCTTCTCT CTGTATATAG
2401 AGAATAAGT TATTGGCCGC GCGCTCTCCT TCAGTCCAGG GTACTACCGG
```

WO 01/12659

PCT/IB00/01496

```
2451 GGCCTCCGCT CGTCCCTCTT CTAGTGGTAC CGCCAGGCC TTAATCACCC
2501 CCATTCCGCTG CGTCCGTATC TCCAGGCTC TACATTCTCG GGAGCGCGCG
2551 CTCCCAAGGG GGTCCCTGGA CTTCTCGCG CTCTCTCTGG CCTCTGAGGG
2601 ATGCGTCTCTA CCGCGCCAT CCGCCCTGG CCGAGAGGG GGAGCTCCCG
2651 TTAGTCCGTC CTCCACCGC CGGCGCTCG CCGCATCCG GGCCTTATCG
2701 ACTGCCCTTC CCACCGGCC CGGCGAGGC AGCGCCGACC CGGCGCGGGG
2751 CACCGCCGAC CGAGCCATCC TGGCTCGCT CCGCCGAGC CTGCAGCTTC
2801 TCGCGAGGGG CGGCGAGGT CCGCTGGTG CAGGAGGGGG TCCCTCTGTT
2851 CGGGGTGAGG CGCTGCTCT CTATTTCAG ATGTTGCTGT AGAATPAAG
2901 ACGGTTTAAA TCTGAAAAA AAAAAAAAA AAAAAAAAA AAAAAAAAA
2951 AAAAAAAAA AAAAAAAAA AAAAAAAAA AAAAAAAAA AA
```

#### BLAST Results

No BLAST result

#### Medline entries

9828268:  
Two kinesin light chain genes in mice. Identification and  
characterization of the encoded proteins.

#### Peptide information for frame 3

ORF from 144 bp to 2009 bp; peptide length: 622  
Category: strong similarity to known protein  
Prosite motifs: RGD (502-505)  
KINESIN\_LIGHT (223-265)  
KINESIN\_LIGHT (265-307)

```
1 MANMVFREE KLSQDEIVLG TKAVIQGLET LRGEHRALLA PLVAPEAGEA
51 EPGSQERCIL LRSLEAIEL GLGEAQVILA LSSHLAGVES EKQKLAQVR
101 RLVOENQWLR EELAGTQOKL QRSEQAVAGL EEEKQHLIFM SQIRKLDEDA
151 SPNEEKGDVP KDTLDDLFPN EDEQSPAPSP GGGDVSGQHG GVEIPARLRT
201 LHMVVIQYAS QGRYEVAVPL CQALEDLER TSGMDPDAV TMLNIALVY
251 RQHWYKEAA HLLNDALAIR ERTLCKDMPA VAATLNLAV LYGRGRVKE
301 AEPLCKRALE IREKVLGKFH POVAKQLSNL ALLCONQKRA EEEVYYRRA
351 LEIYATRLGP DDPNVARTKN NLASCYLKQG KYQDAETLYK EILTRAHEKE
401 FGSVNGDNKP IWMHAEREE SKDKRRDSAP YGEYGSWYKA CKVDSPVTNT
451 TIASLGALYR ROKLEAANT LEDCASRRRQ QGLPASQTK VVELLDGSG
501 RRGDRRSSRD NAGGACPRSE SDLEDVCPFA EWMGDGSGSL RRSQSFGLR
551 DALRRSSEML VKKLQGGTPO EPPNPRMKRA SSLNFWKSV BEPTQPGGTG
601 LSDSRTLSSS SMDLSRRSSL VG
```

#### BLASTP hits

No BLASTP hits available

#### Alert BLASTP hits for DKFZphtes3\_4h6, frame 3

TREMBL:AF055666.1 gene: "Klc2"; product: "kinesin light chain 2"; Mus  
musculus kinesin light chain 2 (Klc2) mRNA, complete cds., N = 1, Score  
= 2824, P = 4e-294

PIR:I53013 kinesin light chain - human, N = 1, Score = 1927, P =  
4.5e-199

PIR:C41339 kinesin light chain C - rat, N = 1, Score = 1919, P =  
3.2e-198

SWISSPROT:KNLC\_RAT KINESIN LIGHT CHAIN (KLC), N = 1, Score = 1919, P =  
3.2e-198

>TREMBL:AF055666.1 gene: "Klc2"; product: "kinesin light chain 2"; Mus  
musculus kinesin light chain 2 (Klc2) mRNA, complete cds.  
Length = 599

HSPs:

Score = 2824 (423.7 bits), Expect = 4.0e-294, P = 4.0e-294  
Identities = 558/598 (93%), Positives = 572/598 (95%)

```
Query: 1 NAMNVFPREKLSQDEIVLGTAVIQGLETLAGEHRAALLAPLVAPEAGEAEPGSGERCIL 60
NA NV PREKLSQDEIVLGTAVIQGLETLAGEHRAALLAPL + EAGEAEPGSGERC+L
Sbjct: 1 NATHVLPREKLSQDEIVLGTAVIQGLETLAGEHRAALLAPLASHEAGEAEPGSGERCIL 60

Query: 61 LRRSLEATELGLGEAQVILALSSHLGAVESEKQKLRQVRLVQENQWLREELAGTQOKL 120
LRRSLEATELGLGEAQVILALSSHLGAVESEKQKLRQVRLVQENQWLREELAGTQOKL
Sbjct: 61 LRRSLEATELGLGEAQVILALSSHLGAVESEKQKLRQVRLVQENQWLREELAGTQOKL 120

Query: 121 QRSEQVAQLEEEKQHLFMSQIRKLDEDAFPNEEKGDVPKDTLDDLPNEDEQSPAPSP 180
QRSEQVAQLEEEKQHLFMSQIRKLDE P EEKGDVPKD+LDOLFPNEDEQSPAPSP
Sbjct: 121 QRSEQVAQLEEEKQHLFMSQIRKLDE-KLPQEEKGDVPKDSLDOLFPNEDEQSPAPSP 179

Query: 181 GGGDVSGQHGGEI PARLRTLHNLVIQYASQGRYEVAVPLCKQALEDEKTSGHDPDVA 240
GGGDV+ QHGGEI PARLRTLHNLVIQYASQGRYEVAVPLCKQALEDEKTSGHDPDVA
Sbjct: 180 GGGDVAAQHGGEI PARLRTLHNLVIQYASQGRYEVAVPLCKQALEDEKTSGHDPDVA 239

Query: 241 TMLNIALVYRDQNKYKAAHLLNDALAIREKTLGKDPAAVATLNNLAVLYGKRGYKE 300
TMLNIALVYRDQNKYK+AAHLLNDALAIREKTLGKDPAAVATLNNLAVLYGKRGYKE
Sbjct: 240 TMLNIALVYRDQNKYKAAHLLNDALAIREKTLGKDPAAVATLNNLAVLYGKRGYKE 299

Query: 301 AEPLCKRALEIREKVLGKFPDVAQKLSMLALLCQNGKAEVEYYRRALEIYATRLGP 360
AEPLCKRALEIREKVLGKFPDVAQKLSMLALLCQNGKAEVEYYRRALEIYATRLGP
Sbjct: 300 AEPLCKRALEIREKVLGKFPDVAQKLSMLALLCQNGKAEVEYYRRALEIYATRLGP 359

Query: 361 DDPNVAKTKNNLASCYLKQGYQDAETLYKEILTRAHEKEFGSVNGKPIWMAEERE 420
DDPNVAKTKNNLASCYLKQGYQDAETLYKEILTRAHEKEFGSVNG+KPIWMAEERE
Sbjct: 360 DDPNVAKTKNNLASCYLKQGYQDAETLYKEILTRAHEKEFGSVNGKPIWMAEERE 419

Query: 421 SKDKRRDSAPYEGVSGMYKACKVDSPTVNTTLASLGALYARQKLEAANTLEDCASTRK 480
SKDKRRD P EYSGMYKACKVDSPTVNTTL+LGLALYR+GLEAANTLEDCASTRK
Sbjct: 420 SKDKRRDRPM-EYSGMYKACKVDSPTVNTTLATLGLALYRPEGKLEAANTLEDCASTRK 478

Query: 481 QGLDPASQTKVVELLKDGSGRRSSRDMAAGGAPRSESDLEDVGPTAEWNGDGSGL 540
QGLDPASQTKVVELLKDGSGRRSSRDMAAGGAPRSESDLEDVGPTAEWNGDGSGL
Sbjct: 479 QGLDPASQTKVVELLKDGSGR-GHRRGSRDVG---POSESDLEESGPAEWSGDGSGL 534

Query: 541 RRSGSFGKLRDALRRSSEMLVKLQGGTPOEPPNPMKRASSLWFLNKSVEEPTQPGG 598
RRSGSFGKLRDALRRSSEMLV+KLQGG PQEP N RMRASSLWFLNKSVEEP QPGG
Sbjct: 535 RRSGSFGKLRDALRRSSEMLVRKLQGGQPQEP-NRMRASSLWFLNKSVEEPVQPGG 591
```

Pedant information for DKFZphtes3\_4h6, frame 3

Report for DKFZphtes3\_4h6.3

```
(LENGTH) 622
(MW) 68934.82
(pI) 6.72
(HOMOL) TREMBL:AF055665.1 gene: "Klc2"; product: "kinesin light chain 2"; Mus musculus
Kinesin light chain 2 (Klc2) mRNA, complete cds. 0.0
(BLOCKS) BL00927C Trehalase proteins
(BLOCKS) BL01160I Kinesin light chain repeat proteins
(BLOCKS) BL01160H Kinesin light chain repeat proteins
(BLOCKS) BL01160C Kinesin light chain repeat proteins
(BLOCKS) BL01160F Kinesin light chain repeat proteins
(BLOCKS) BL01160E Kinesin light chain repeat proteins
(BLOCKS) BL01160D Kinesin light chain repeat proteins
(BLOCKS) BL01160C Kinesin light chain repeat proteins
(BLOCKS) BL01160H Kinesin light chain repeat proteins
(BLOCKS) BL01160A Kinesin light chain repeat proteins
(SUPFAM) tetratricopeptide repeat homology la-07
(PROSITE) RGD 1
(PROSITE) HYRISTYL 8
(PROSITE) KINESIN_LIGHT 2
(PROSITE) AMIDATION 2
(PROSITE) CAMP_PHOSPHO_SITE 5
(PROSITE) CK2_PHOSPHO_SITE 11
(PROSITE) TYR_PHOSPHO_SITE 3
(PROSITE) PKC_PHOSPHO_SITE 7
(PROSITE) ASN_GLYCOSYLATION 2
(PFAM) Kinesin light chain repeat
(KW) All Alpha
(KW) LOW_COMPLEXITY 12.54 %
(KW) COILED_COIL 4.98 %
```

```
SEQ MAMVFPREKLSQDEIVLGTAKVIQGLTIRGEHALLAPLVAPEAGEEPGSGERCIL
SEG .....
PRD ccccccccccccccccccccccccccccccccccccccccccccccccccccccccc
COILS .....

SEQ LRSLLEAIELGLGEAOVILALSSHLGAVESEKOKLRAQVRLVOENQWLRELAGTOOKL
SEG .....
PRD hhhhhhhhhcccccccccccccccccccccccccccccccccccccccccccccccc
COILS .....CCCCCCCCCCCC

SEQ QRSQAVAQLEEEKQHLFMGQIRKLDEASPNEEKGQVPKDTLDDLFPNEDEQS PAPS P
SEG .....
PRD hhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh
COILS CCCCCCCCCCCCCCCCCC.....

SEQ GGGVSGQHGQYEI PARLRTLHNLVIQYASQGRYEVAVPLCKQALEDEKTS GHDHPOVA
SEG .....
PRD ccccccccccccccccccccccccccccccccccccccccccccccccccccccccc
COILS .....

SEQ TMLNIALVYRDQNKYREAAHLNDALIREKT LGKH PAVAAATLNNLAVLYGRGRGYKE
SEG .....XXXXXXXXXXXX
PRD hhhhhhhhhhhcccccccccccccccccccccccccccccccccccccccccccccccc
COILS .....

SEQ AEFCKRALEIREKVLGKFHPDQAKQLSHLALLCQMGKAEVEYYRRALEIYATRLGP
SEG .....
PRD hhhhhhhhhhhhhcccccccccccccccccccccccccccccccccccccccccccccc
COILS .....

SEQ DDPNVAKTKNNLASCYLKQGRYQDAETLYKEILTRAHEKEFGSVNGDNKPIWMAEEREE
SEG .....XXXXXXXX
PRD ccccccccccccccccccccccccccccccccccccccccccccccccccccccccc
COILS .....

SEQ SKDKRRDSAPYGEYSWYKACKVDSPTVNTLRLS LGALYRQGRKLEAHTLEDCASRWK
SEG .....XXXXXXXX
PRD hhhhhcccccccccccccccccccccccccccccccccccccccccccccccccccccc
COILS .....

SEQ QGLDPASQTKVVELLKDGSGRRGDRSSRD MAGGAGPRSESDLEDVGPTAEWNGDGSGL
SEG .....XXXXXXXXXXXX
PRD hhcccccccccccccccccccccccccccccccccccccccccccccccccccccccc
COILS .....

SEQ RRSQSPKLRDALRRSSEMLVKKLGGGT PQEPHPKRRASSLWFLNKSVEEPTOPGGTG
SEG .....XXXXXXXXXXXX
PRD ccccccccccccccccccccccccccccccccccccccccccccccccccccccccc
COILS .....

SEQ LSSDRTLSSSMDLSRRSLVG
SEG .....XXXXXXXXXXXX
PRD ccccccccccccccccccccccc
COILS .....
```

## Prosite for DKF2phtes3\_4h6.3

PS000001	449->453	ASN_GLYCOSYLATION	PD0C00001
PS000001	587->591	ASN_GLYCOSYLATION	PD0C00001
PS000004	425->429	CAMP_PHOSPHO_SITE	PD0C00004
PS000004	505->509	CAMP_PHOSPHO_SITE	PD0C00004
PS000004	554->558	CAMP_PHOSPHO_SITE	PD0C00004
PS000004	578->582	CAMP_PHOSPHO_SITE	PD0C00004
PS000004	616->620	CAMP_PHOSPHO_SITE	PD0C00004
PS000005	30->33	PKC_PHOSPHO_SITE	PD0C00005
PS000005	90->93	PKC_PHOSPHO_SITE	PD0C00005
PS000005	451->454	PKC_PHOSPHO_SITE	PD0C00005
PS000005	499->502	PKC_PHOSPHO_SITE	PD0C00005
PS000005	507->510	PKC_PHOSPHO_SITE	PD0C00005
PS000005	539->542	PKC_PHOSPHO_SITE	PD0C00005
PS000005	615->618	PKC_PHOSPHO_SITE	PD0C00005
PS000006	13->17	CK2_PHOSPHO_SITE	PD0C00006
PS000006	151->155	CK2_PHOSPHO_SITE	PD0C00006
PS000006	163->167	CK2_PHOSPHO_SITE	PD0C00006
PS000006	232->236	CK2_PHOSPHO_SITE	PD0C00006
PS000006	470->474	CK2_PHOSPHO_SITE	PD0C00006
PS000006	507->511	CK2_PHOSPHO_SITE	PD0C00006
PS000006	519->523	CK2_PHOSPHO_SITE	PD0C00006
PS000006	521->525	CK2_PHOSPHO_SITE	PD0C00006



PS00006	568->572	CK2_PHOSPHO_SITE	PDOC00006
PS00006	589->593	CK2_PHOSPHO_SITE	PDOC00006
PS00006	610->614	CK2_PHOSPHO_SITE	PDOC00006
PS00007	339->346	TYR_PHOSPHO_SITE	PDOC00007
PS00007	339->347	TYR_PHOSPHO_SITE	PDOC00007
PS00007	424->432	TYR_PHOSPHO_SITE	PDOC00007
PS00008	71->77	MYRISTYL	PDOC00008
PS00008	86->92	MYRISTYL	PDOC00008
PS00008	182->188	MYRISTYL	PDOC00008
PS00008	187->193	MYRISTYL	PDOC00008
PS00008	402->408	MYRISTYL	PDOC00008
PS00008	482->488	MYRISTYL	PDOC00008
PS00008	598->604	MYRISTYL	PDOC00008
PS00008	600->606	MYRISTYL	PDOC00008
PS00009	292->296	AMIDATION	PDOC00009
PS00009	499->503	AMIDATION	PDOC00009
PS00016	502->505	RGD	PDOC00016
PS01160	223->265	KINESIN_LIGHT	PDOC00893
PS01160	265->307	KINESIN_LIGHT	PDOC00893

## Pfam for Dkf2phtes3\_4h6.3

HMM\_NAME Kinesin light chain repeat  
 HMM \*RALEDREKTLGHDHPDVAIHLNMLALVCRNQNKYeEveNYTN\*  
 \*ALED\*EKT\*GHDHPDVAIHLN\*LAIV\*E\*QNKY\*E++ ++N  
 Query 223 QALEDEKTSQHDHPDVAIHLNMLALVYRDQNKYKEAAHLLN 264  
 50.46 265 306 1 42 dkf2phtes3\_4h6.3 strong similarity to Kinesin light chain  
 Alignment to HMM consensus:  
 Query \*RALEDREKTLGHDHPDVAIHLNMLALVCRNQNKYeEveNYTN\*  
 AL\*REKTLG DHP VA LNNLA+++ ++KY\*E\*E + +  
 dkf2phtes3 265 DALAIREKTLGKDHFAVAATLNNLAIVLGVKRGKYKEAEPLCK 306  
 Query 348 1 42 dkf2phtes3\_4h6.3 strong similarity to Kinesin light chain  
 Alignment to HMM consensus:  
 HMM \*RALEDREKTLGHDHPDVAIHLNMLALVCRNQNKYeEveNYTN\*  
 RALE\*REK\*LG HPDVA++L\*NLAL\*G\*QNK\*EEVE YY\*  
 Query 307 RALEIREKVLGHRFPDVAKQLSNLALLCQNGRAEEVEYYR 348  
 39.10 349 390 1 42 dkf2phtes3\_4h6.3 strong similarity to Kinesin light chain  
 Alignment to HMM consensus:  
 Query \*RALEDREKTLGHDHPDVAIHLNMLALVCRNQNKYeEveNYTN\*  
 RALE+ LG D P\*VA+ NNLA + Q\*KY+++E +Y+  
 dkf2phtes3 349 RALEIYATRLGPDHPVAKTKNNLASCYLKQGRYQDAETLYK 390

WO 01/12659

PCT/IB00/01496

DKFZphes3\_4o19  
-----

group: testes derived

DKFZphes3\_4o19 encodes a novel 1180 amino acid protein with weak similarity to human megakaryocyte stimulating factor and human mucin.

The novel protein contains a cytochrome c family heme-binding site signature.  
No informative BLAST results; No predictive prosite, pfam or SCOP motif.

The new protein can find application in studying the expression profile of testis-specific genes.

similarity to megakaryocyte stimulating factor and mucin

complete cDNA, complete cds, EST hits (few)

Sequenced by AGOWA

Locus: unknown

Insert length: 3767 bp

Poly A stretch at pos. 3757, polyadenylation signal at pos. 3737

```

1 GGCTAGGTTT AGCTTCAGGG GCAGCCAGG GCAGTGTTC TGCATATTGC
51 ATGGATGAAA GGCCTAAGGC TGCCCTCTCT TGCAGGCTGG CTTCAGAGAT
101 TGCACCTTCT TCTCCTGCTA CTCCTCCAAA TCTATGACCC TTCAAGGCAG
151 AGCTGACCTG TCGGTAATC AAGGCAATGC AGCCGGCCGC CTAGCTACAG
201 TTACAGAGGC AGTGTGACG CAGTGGGGGG TGCACTCTCC AGCCCGGCT
251 CACCCAGTTC TCTGGACAA AATGGAGAAA GCGCTCCAC AGCCCAAGCA
301 CGAGGCGCTC AAGTCCAAGG AGCATCTTCC GCAACAGCCT GCCGAGGCA
351 AGACGGCGTC CCGCGCGCTC CCAGCGCTCC GGGCTGTGT AGAGAGCCAG
401 GCTTTCAGGA ACATCTCGGT AGACGAGATG GACATGATC AGCGCGGTGC
451 AGCCAGCGTC ATCCAAGCCA ATGGAGGGGG CTATTGGCTC CGGCAAGGC
501 TGATTTCCCA GATGATGGCG GCCAAGGCCA TCCAGAGGC CTGGCGGCGC
551 TTCAACAAGA GACACATCCT TCACTCCAGC AAGTGTGTGG TAAAGAAAAC
601 GAGGGCGGAG GAGGGGGACA TACCTTATCA CCCCACAGC CAGGTGGCT
651 TCCAGCATCC GGAAGAGAAC CGCTTCTGT CCCCGCCAT CATGCTGAAC
701 AAGGAGACCC AGTTCCCTTC CTGTGACAAAT CTGGTCTCT GCAGACCCCA
751 GTCTGTCCTC CTCTGCGAG CCCCAGCAGC TCAGGGTACC CCAGAGCCCT
801 GTGTGACAGG TCTCATGCTT GGCAGAGTCC GGGGGTGGC CTTCCTGCCA
851 CACCAAGCGG TCCCATCAG ATTTCCCTGC CCAGTGAATT TGGAGCCAAA
901 ATGCCAGCCA TGCTGTGTA CCAGAACCAT CAGAAGCACC TGCTCTGCTC
951 ACATAGAGGG TGACTCACTG AAGACCAACG GTGTAGTGC CCGGACCAAC
1001 AAGCCAGGGG CTCGGAGAGC ACCATTGTCC AGAAGGTATG ACCAGGCACT
1051 TACGAGACCA TCCAGAGCC AAGCCAGGG CCTGTGAAG GCAAGAGCCC
1101 CCAAGACCCC CTTCAGATA TGTCAGGGCG CCATGATCAC CAAGACTCTA
1151 TCCAGACAT ATCCAGTGT CTCTGTGACC CTGCCACAGA CATATCCAGC
1201 GTCCACGATG ACCAGCACCC CACCCAGAGC TAGCCAGTT CCCAAGTAA
1251 CAATATCAAA GACCCACCCC CAGATGTAT CCGGGCCAC AGTGACCAAA
1301 ACTGCACTTC ACACATGCC CATGCCACA ATGACCAAGA TCCAGGTACA
1351 CCCACAGCC TCCAGAACTG GCACCCACGC GCAGACATGC CTTGCCACCA
1401 TCACGGGAAA GAACCGACCT CAGGTTTCCC TTCTGGCTTC CATCATGAAG
1451 AGCTTGCCTC AGGTATGCC GGGGCTGGG ATGCCAAGA CCCCACCCA
1501 GATGCACCGG GTACACACCC CAGCCAAAAA CCAATGCAA ACATGTCTGT
1551 CAGGCCAAT GTCCAAGACT TCATCCCAAG GGAGCCAGT TGGGTGACC
1601 AAGCCCTCAC CCGAGACCCG CCTGCCAGCC ATGATAACCA AGACCCACAG
1651 CAGTTACGCT TGGGTGGCCA CATCTCTAA GACTCTGTGT CTGGCTCTC
1701 CAACAGTGGC AATGTCAAAG GCTCCACCCC AAGTGGCGT AGCAGCCGGA
1751 ACTCCAAACA CTTCAAGGTC CATCATGAG AACCCACCA AGGCCAAGGC
1801 CACCGTAAAT GTGAAGCAG CTGCCAAGGT GGTGAAGGC TCATCCCCCT
1851 CTTATTTTGG TGAGGGGAG ATCAGGTGCC TGCTCAAGC ACATCCGGGA
1901 ACTGGGGTCC CAGGGCTGC AGCTGAGCT CTTTTGAAG CCGAGAAAT
1951 CAAGACTGGC ACCCAGAAC AGCGAAAAAC AGCATGGCA TTTAAGACCA
2001 GTGTGGCAGT GGAATGGCT GGGGCTCCAT CCTGGACAAA AGTTGCTGAG
2051 GAAGGGGACA AGCCACCTCA CGTGTATGT CCTGTAGACA TGGCTGTGAC
2101 CTGCCCCCGG GAGACCTGT CTGCCCATC GACCAATGCC TCATCCGAGA
2151 GACATCCACC CTGCTGTCC CAGAGACCA TGCCCGCCCC GCTGACCAAG
2201 GCTCATCTC AGGACATCT GCCCACTGAG CTGACCAAGA CCCCATCCCT
2251 GGCCCATCTG GACACCTGTC TGAGCAAGAT GCATTCCAG ACACATCTGG
2301 CCACAGGTGC CTGAAGGTC CAGTCCCAAG CGCTCTAGC CACTGTCTG
2351 ACCAAGAGGC AGTCCGGGZ GCAAGCGATC ACAGACATAA CCGCTGCTT
2401 CATCCACGCG CACCAAGCTG CTGATCTCAG CAGCAACACC CACTCCAGG
2451 TGCTCTAAC AGGGTCCAA GTGTCCAACC ACGCCTGCCA GCGCCTCGGT
2501 GGCCTCAGCG CCCCACCTG GGGCAAGCCA GAGGACAGAC AGACCAAGCC
2551 AGAGCCGAC GACACGTGC GGGGAGAGC CACTCAAGG GACCATGCC
2601 CGCAGCGCTG TGAGTCCAG GGTATGCTG TGCCGCCAT GGCACCCACC

```

TREMBL:HSN2A.1 gene: "MUC2"; product: "mucin"; Human mucin-2 gene, partial cds., N = 1, Score = 204, P = 1.4e-12  
PIR:S48478 glucan 1,4-alpha-glucosidase (EC 3.2.1.3) - yeast (Saccharomyces cerevisiae), N = 1, Score = 192, P = 9.6e-11

>TREMBL:HSU70136.1 product: "megakaryocyte stimulating factor"; Human megakaryocyte stimulating factor mRNA, complete cds.  
Length = 1,404

## HSPs:

Score = 242 (36.3 bits), Expect = 9.6e-16, Sum P(2) = 9.6e-16  
Identities = 145/546 (26%), Positives = 198/546 (36%)

Query: 282 KRVSARTNKARAPETPLSRRYDQAVTRPSRAQTGQGVKAEETPKAPFQIC-PGPMITKLL 340  
K+ + T K AP TP PS + P T AP P P TK+  
Sbjct: 488 KKPAPITPKAPAPITPKAPAPITPKAPAPITPKAPAPITPKAPAPITPKAPAPITPKAP 546  
Query: 341 QTPVSVSVTLPO---TYFASNTMTTPPKTSV-PKVITIKTPAQMYPGPTVTKTAPHTC 395  
T S T + T P TTP K + P PK TP + P PT TK  
Sbjct: 547 TTPKEPSPTTKEAPAPITPKAPAPITPKAPAPITPKAPAPITPKAPAPITPKAPAPITPK 599  
Query: 396 PMPTMTKIQVHTASRTGTPTCTPATITAKNRQVSLASIMKSLPOVCPGPAKATPP 455  
P PT K + PT TP++T P T LA P +A T P  
Sbjct: 600 PAPTAPK-EPAPT-----TPKETAPTTPKKLTPTTPEKLAPITPEKAPAPITPEELAPITP 653  
Query: 456 QMHPVTTTAKNPLOTCLSATMSKTSSQSPVGVTKPSPOT-RLPAMIT-KTPAQLRSVAT 513  
+ TTP + P T A T + +P +P+P T + PA T K A T  
Sbjct: 654 EEPPTTTP-EEPAPITPKAAAPNTPKEAPAPITPKAPAPITPKAPAPITPKAPAPITPKGT 712  
Query: 514 ILKTLCLASPTVANVKAPQVAVAG---TPMTSGSINENPPKAKATVNVQAAKVV-KA 569  
TL +PT AP +A T TS PK A K+ A K  
Sbjct: 713 APTTLKEAPAPITPKKPAKELAPITTKPTSTSTSDKAPITPKGTAPITPKAPAPITPK 772  
Query: 570 SSPSYLAEGKIRCLAQHPGTGVPRAAELPLEAEIKTGT--QKQAKTMAFKTSVAVE 627  
+P+ L +P P T A EL K T T K A T +T+  
Sbjct: 773 PAPTTPKGTAPITLKEAPAPITPKKPAKELAPITTKGTSTSTSDKAPITPK-ETAPITP 831  
Query: 628 MAGAPSWTKVAEGDGKPHVYVVDMAVTLPRGQALAPLWASSQRPCLSORPLAAPL 687  
AP+ K + P P P V+ P +S P LS P L  
Sbjct: 832 KEPAPITPK--KPAPITPETPPPTSEVSTPTTTKEPTTIHKSPOESTPELSAEPTRKAL 889  
Query: 688 TKASSQGHLPTELKTPSLA--HLOTCLSKHISOTHLATGAVKVOSQAPLAT--CLTKTQ 743  
+ + +PT TKTP+ + T ++ L T + + AP T T T+  
Sbjct: 890 ENSPKPEGVPT--TKTPAATKPEMTTAKDKTTERDLRT-TPETTAAKPKTKETATTTE 946  
Query: 744 SRGQPTIDITCLIFANGAADLS--SHTSHQVLLTGSKVSH--HACQRLGGLSAPP-MAK 798  
+ TT + + D + T + KV+ ++ P AK  
Sbjct: 947 KTIESKITATTTQVTSITTTQOTTFFKILTLLKTTLPKVTITTKTITTEINNKPEETAK 1006  
Query: 799 PEDROTQOPHGHVPGKTTQGGPCPAA 825  
P+DR T + P K T+ P +  
Sbjct: 1007 PKDRATNKATTPKPKQKPKAKPKPTS 1033  
Score = 205 (30.8 bits), Expect = 3.1e-12, Sum P(2) = 3.1e-12  
Identities = 146/565 (25%), Positives = 209/565 (36%)  
Query: 281 KRVSARTNKARAPETPLSRRYDQAVTRPSRAQTGQGVKAE--TPKAPFQICPGPMITKT 338  
TK+ + K AP TP + A T P + P K TP+ P D + T  
Sbjct: 597 TKKAPAPAPKEAPAPITPK----ETAPITPKKLTPTTPEKLAPITPEKAPAPITPEELAPIT 652  
Query: 339 LLQTPVSVSVTLPOQTPASTNTTTPPKTSV-PKVITIKTPAQMYPGPTVTK-TAPHTCP 396  
+ P T P + TP + +P PK TP + P PT K TAP T P  
Sbjct: 653 PEEPTTTPPEAPAPITPKAAAPNTPKEAPAPITPKAPAPITPKAPAPITPKAPAPITPK 709  
Query: 397 N---PTMTKIQVHTASRTGTPTCTPATITAKNRQVSLASIMKSLPOVCPGPAKAT 453  
PT K + PT + P+ P T + S + K P G A T  
Sbjct: 710 KGTAPITLK-EPAPITPKKPAKELAPIT---TKPTSTSD--KPAPITPKGTAPT-T 761  
Query: 454 PPQMHVTTTAKNPLOTCLSATMSKTSSQSPVGVTKPSPOTRLPAMITKTTPAQLRSVAT 513  
P + P TTP K P T T T + +P KP+P+ P TK P S  
Sbjct: 762 PKAPAP-TTP-KEAPAPITPKGTAPITLKEAPAPITPKKPAKELAPIT-TKGTSTSTSDK 818  
Query: 514 ILKTLCLASPTVANVKAPQVAVAGTPTSGSINENPPKAKATVNV---KQAAKVVKA 569  
T +PT AP A P T E PP + V+ K+ + K+  
Sbjct: 819 APTTPKETAPITPKAPAPITPKKPA--PTTP----ETPPPTTSEVSTPTTTKEPTTIHKS 872  
Query: 570 ---SSPSYLAEGKIRCLAQHPGTGVPRAAELPLEAEIKTGTQKQAKTMAFKTSVAV 626  
S+P AE + L GVP + P + T T K T+ +T+

WO 01/42659

PCT/IB00/01496

```
2651 GGCCATTCCA CATGCAACGT TGAGTCCTGG GGAGACAACG GAGCCACACG
2701 TGCCGAGCEA TCAATGCCCG GCGAGGCGGT GCCCTGCCAG GAGGACACGG
2751 GCGCCGGGGA CGCTGGTGTG GTTGTGGGCT AATCGTGAA CCGCGCATGG
2801 GAGCCAGCEA GGGGTGCTGC GTCTGGGAC ACCTGGGCA ACAAGGCGGT
2851 GGTGCTCCCG AGGCGGTCCG GGGAGCCAAT GGTGTCCATG CAGGCTGCAG
2901 AGGAGATCCG CATCTTCGCA GTGATCACTA TCCAGGCGGG CGTCCGTGGC
2951 TACCTGGGCG GTGCGAGGAT CCGGCTGTGG CACGGGGGGG CCATGGTCAT
3001 CCAAGCTACT TGGCGCGGCT ACCGTGTGGG GCGGAACCTG GCACACCTCT
3051 CGAGAGCCAC CAGGACCATC CAGTCTGCCT GGGCGGGCTA CAGCACCCGC
3101 CGGGACCAAG CCGGCACTG GCAGATGCTC CACCCCTCA CGTGGGTGGA
3151 GCTGGGCGAG CGGGCGGGG TCATGCTGTA CGGAACCTG TTCCAGATG
3201 CGAGAGCCAG GACAGTATCT GACCATCGCT GCTTCCAGTC CTGCGAGGCA
3251 CAGGCTTGCA GCGTCTGCCA CTCCTGAGC TCCAGGATCG GGAGCCGCC
3301 CAGCGTGTGT ATGCTAGTGG GCTCCAGGCC TCGCACCTGT CATACCTGTG
3351 GAGCCACACA GCCCACCCTG GTGGTGCAGG GCATGGGCA GGGCAGTAC
3401 GCGCCGGGGG CAGTGTCTTG GGGCTTGGG TACGAGCTGG CTGGCTCTAG
3451 TCCGAGGCGG CCGCATCGCC AGGACAAGC GGGCAGGCGC ATCCAGTCCG
3501 CCTGGAGGGG CTTTAAGATC CGCCAGCAGA TGAGGAGCGA GCAATGGCA
3551 GCGAGATAG TTCAAGCCAC CTGGGAGGCG CACCATACCC GGAGCTGTCT
3601 GAGGACACA GAGGCGCTCT TGGAGCAGG AGACCCCTGG GCGAGTCA
3651 GGCACATGCA TTGGCCTGGC ATCTAGGACC CTGGCTCCCT GCAGTGGGGA
3701 CTTCGTGGGA GGCACCTATG GCTCTCTGGG TCTAATGAAT AAAGTCTCTC
3751 ACAGCCTAAA AAAAAAA
```

BLAST Results

No BLAST result

Medline entries

No Medline entry

Peptide information for frame 2

ORF from 134 bp to 3673 bp; peptide length: 1180  
Category: similarity to known protein

```
1  MTLQGRADLS  GNOGNAAGRL  ATVHEPVVTO  WNVHPPAPAH  PSLLDKMEKA
51  PPQPQHEGLK  SKEHLPOOPA  EGKTASRRVP  RLRAVVESEA  FKNILVDEMO
101 MMHARAATLI  QANWRGYWLR  QKLSIQHMA  KAIGEAHRRF  NKRHILHSSK
151 SLVKRTAEAE  GDIPIYAPQ  VRFQHPEDH  LLSPIHMKW  EYQFPGCDL
201 VLCRPOSSPL  LQPPAAGTPT  EPCVQPHAA  RVRLAFLPH  OTVIRFPCL
251 VSLDAKQPC  LLTRTIRSTC  LVHIEGDSVK  TKRVASRTNK  ARAPETPLSR
301 RYQAVTRPS  RAQTQGVVKA  ETPKAPFQIC  PGPHITKILL  QTPVVSUTL
351 PPTYPASTNT  TTPKTSVVP  KVILIKTQ  MYGPPVTKT  APHICMPTM
401 TKIQVHTAS  RTCTPRQTC  ATITAKNRPO  VSLIASINKS  LPQVCPGPM
451 AKTPPQMHVP  TTPAKNFLQT  CLSATMSKTS  SQSPVGVTK  PSQTRLFPM
501 ITRTPAQLRS  VATILKTLCL  ASPTVAHVYA  PPQVAVAGT  PNTSGSINEN
551 PPKAKATVNV  KQAAKVVEAS  SPSVLAEGKI  RCLAQHPHT  GVPRAAEELP
601 LEAEIKTKCT  OKQAKTDMAF  KTSVAVENAG  APSWTKVAEE  GQKPPHYVVP
651 VMAVTLPRG  OLAAPLTNAS  SQRHPPCLSG  RPLAAPTKA  SSQHLPTL
701 TKTPSLAHL  TCLSRHMSQT  HLTGAVKVQ  SQAPLATCLT  KTQSRGQFIT
751 DITTCILPAM  QAADLSNTH  SOVLLTGSK  SHACORLGG  LSAPPMWPE
801 DRGTQOPHG  HVPGTITGCG  PCPACCEVCG  MLVPPMAFTG  HSTCNVESWG
851 DNGATRAQPS  MFGQAVPCQE  DTCPADAGVV  GGOSWNAME  PARGAASWT
901 WRNKAVVFP  RSGEPVMSQ  AAETIRILAV  ITIQAGVRGY  LARRIRLWH
951 RGAHVIGATM  RGVHVRNLA  HLCRATTTIG  SAWRGSTRR  DQARHWOHLH
1001 PTHVELGSR  AGVMSDSWF  CQCRAATVSD  HRCSTQCOAH  ACSVCHSLSS
1051 RIGSPSPVVM  LVGSSPRTCH  TCGRTQPTRV  VQHGQGTGEC  PCAVSWASAY
1101 OLAALSPPRP  HRQDKAATAI  QSAWRGFKIR  QMRQOQMAA  KIVQATWRGH
1151 HTRSCLEKTE  ALLGPADPSA  SSRHMHWFGI
```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DNF2phtes3\_4ol9, frame 2

TREMBL:HSU70136.1 product: "megakaryocyte stimulating factor"; Human  
megakaryocyte stimulating factor mRNA, complete cds., N = 2, Score =  
242, E = 9.6e-16

Sbjct: 873 PDESTPELSAEPKALENSKEPGVP--TTKTPATKPEMTTAKDKTTERDLATTET 930

Query: 627 EMAGAPSMTR-VAEEGDKPPHYVVDNAVTLPRGQLAAPLTNASSQRNPPCLSORPLAA 685  
A AP TK A +K +T Q+ + T + L LA

Sbjct: 931 TTA-APKMTKETATITTEKT-----TESKITATTQVSTTTQDTPFKITTLKTTTLAP 983

Query: 686 PLTKASSOGHLPTLTKTPSLAHLDTCLSKMHSOTHLATGAVKVS-----QAPLATCLT 740  
+T + + TE+ P +T K + AT K Q + P +T

Sbjct: 984 KVT-TTKKTIITTEIMNKPE----ETAKPKORATNSKAT-TPKPKPTKAPKPTSTKKP 1037

Query: 741 KTQSR-GQPTTDT-----TCLIPAHQAADLSNTHSOVLLTGSKVSNHACQRLGGLSAPP 795  
KT R +P T T T + P + Q + + H + S

Sbjct: 1038 KTMFRVRKPKTTTPRKMSTSNPELNPTSRIAEAMLQTTTRPNQTPHSLVEVNPXSEDA 1097

Query: 796 M-AKPEDQTOPQPHGVPGKTTGGPCPAACEVOGMLVPPHAPTHGSHCN 845  
A+ E +PH +P T P OG+++ PH + CN

Sbjct: 1098 GGAEGETHMMLLAPHVHPEVTPMDYLPVRVP-QGIIINHLSDETNIWN 1147

Score = 198 (29.7 bits), Expect = 2.3e-11, Sum P(2) = 2.3e-11  
Identities = 142/513 (27%), Positives = 200/513 (38%)

Query: 204 RPQSSPLLQPPAAQGTPEPCVOGPHAAVRGLAFLPHQTVTIRFPCPVSLDAKCCPLL 263  
R + P +PP G + H V+ + +P L

Sbjct: 207 RTKKKPTPKPPVVDGAGSLONGDFKVTTPDSTTQHNVSTSPKITTAKPINPRPSLPP 266

Query: 264 R--TIRSTCLVHIEGDSVTKRVSARTNKARAP---ETPLSRRYDQAVTRPSR---AQO 315  
T + T L + +V+TK + TNK + E S + Q+ + S A T

Sbjct: 267 NSDTSKETS LTVNKETTETKETT-TNKQTSIDGKEKTSAKETOSIKTSAKOLAPTS 325

Query: 316 GPVKAETPKAPFOICPGPMITKLLQTYPVVSVTLPTQYPASTMTTTPKTSVPVKVII 375  
+ TPKA GP +T T + P T P+ PAST TP + P +

Sbjct: 326 KVLAKPTPKAE--TTKGPALT-TPKEPTP---TPPKE-PAST---TPKEPTTIKSAF 375

Query: 376 KTPAQHYPGPTVTKTAPHTC--PHPTMKIQVHTASRTGTPROT-PATITAKNRQVS 432  
TP + P PT TK+AP T P PT TK + PT + P T PA T K+ P

Sbjct: 376 TTPKE--PAPTTTKSAPTTPEKEPAPTTTK-EPAPTTPEKEPAPTTPEKEPAPTTTKSAPTT 432

Query: 433 ---LLASIMKSLPQVCPGAMAKTPQKHVPVTPAKNPLQCLSATNSKTSQSRSPGV 489  
+ K P FA TP + P TTP K P T + T + +P

Sbjct: 433 KEAPTTTPKKPAPTTPEKEPAPT-TPKEPTP-TTP-KEAPTTPEKEPAPT-TPKEPAPTAPK 488

Query: 490 KPSPT-RLPAMIT-KTPAQLRSVA---TILK---TCLASPTIVANVKAPPOVAVAGT 540  
KP+P T + PA T K PA + T K T ++PT AP A T

Sbjct: 489 KPAPTTPEKEPAPTTPKEPAPTTKEPSPTTPKEPAPTTTKSAPTTPEKEPAPTTTKSAPTT 548

Query: 541 PNT-SGSIHENP----PKAKATVNVQAAKV-KASSPSYLAEGKIRCLAQPHGTGVPR 594  
P S + + P PK A K+ A K +P+ E +P P P+

Sbjct: 549 PKESPSTTTKEPAPTTPEKEPAPTTPKKPAPTTPEKEPAPTTPKEPAPTTTKKAPTA--PK 606

Query: 595 AAAPLEAEKIKTGQKQARTOMAKTSVAHEMAGAPSMTR-VAEEGDKPPHYVVDNA 653  
A P ++ T K+ K + AP+ +A + P P +

Sbjct: 607 EPA--PTTPKETAPTPKLTPPTPEKLAPTTPEKAPTTPEELAPTTPEEPTPTTPEEP 664

Query: 654 AVTLPRGQLAAPLTNASSQRNPPCLSORPLAAPLTASSOGHLPTLTKTPSLAHLDT 712  
A T P+ AAP T + P P + P AP T P E T T

Sbjct: 665 APITPKA--AAPNT----PEKEPAPTPEKEPAPT-APTTPKEPAPTPEKETAPTTPGTAPT 716

Query: 713 LSK 715  
L +

Sbjct: 717 LKE 719

Score = 108 (16.2 bits), Expect = 4.3e-02, Sum P(2) = 4.3e-02  
Identities = 60/214 (28%), Positives = 85/214 (39%)

Query: 265 TIRSTCLVHIEGDSVTKRVSAR-TNKA--RAPETP-LSRRYDQAVTRPSRAQTOGPVKA 320  
T + +H D T +SA T KA +P+ P + A T+P T

Sbjct: 862 TTKEPTTIHKSPDE-STPELSAEPKALENSKEPGVPTTKTPATKPEMTTAKDKTT 920

Query: 321 ETP--KAPFOICPGPMITK-TLLQTYPVVSVTLPTQYPASTMTTTPKTSVPVKVII 377  
E P P +TK T T + T T TTT T+D K+T KT

Sbjct: 921 ERLRTTPETTTAAPKMTKETATITTEKTESKITATTQVSTTTQD-TTFF-KITTLKT 978

Query: 378 PAQHYPGPTVTK---TAPHTCPMPTHT-KIQVHTASRTGTPROTCPATITAKNRQVSL 433  
+ P T TK T P T K + T S+ TP+ P A +P +

Sbjct: 979 TT-LAPKVTITTKTITTEIMNKPEETAKPKORATNSKATTPKPKPTK--APKKPTSTK 1035

Query: 434 LASINKSL--PQVCPGPA-NAKTPQKHVPVTPARNPLOT 470  
H + P+ P P K T P++P + A+ LQT

Sbjct: 1036 KPTMFRVRKPKTTTPRKMSTSNPELNPTSRIAEAMLQ 1075

Score = 56 (8.4 bits), Expect = 3.1e-12, Sum P(2) = 3.1e-12

Identities = 17/60 (28%), Positives = 22/60 (36%)

Query: 22 TVHEPVVTVQAVHPPAPAHPSLLDKMEKAPPOQHEGLKS-KEHLPOQPAEGKTASRRVP 80  
T EP T P P PS E AP P+ + K+ P P E + + P  
Sbjct: 533 TTKEPAPTTTKSAPTTKPEPSPTTKEPAPTTKPEPAPTTKPKPAPTTKPEPAPTTKPEP 592

Score = 52 (7.8 bits), Expect = 9.6e-16, Sum P(2) = 9.6e-16  
Identities = 17/59 (28%), Positives = 22/59 (37%)

Query: 22 TVHEPVVTVQAVHPPAPAHPSLLDKMEKAPPOQHEGLKSKEHLPOQPAE-GKTASRR 78  
T EP T P P P+ E P P+ +KE P P E TA ++  
Sbjct: 431 TPKEPAPTTKPKPAPTTKPEPAPTTKPEPTPTTKEPAPTTKPEPAPTTKPEPAPTTKPKK 489

Score = 51 (7.7 bits), Expect = 1.2e-15, Sum P(2) = 1.2e-15  
Identities = 15/51 (29%), Positives = 19/51 (37%)

Query: 22 TVHEPVVTVQAVHPPAPAHPSLLDKMEKAPPOQHEGLKS-KEHLPOQPAE 71  
T EP T P P P+ + AP P+ + KE P P E  
Sbjct: 416 TTKEPAPTTKSAPTTKPEPAPTTKPKPAPTTKPEPAPTTKPEPTPTTKE 466

Score = 47 (7.1 bits), Expect = 3.2e-15, Sum P(2) = 3.2e-15  
Identities = 12/41 (29%), Positives = 17/41 (41%)

Query: 36 PAPANPSLLDKMEKAPPOQHEGLKSKEHLPOQPAEGKTAS 76  
P P P + P +P +KS P++A T S  
Sbjct: 350 PPTPTTK--EPASTTKPEPTTTIKSAPTTKPEPAPTTKS 388

Score = 47 (7.1 bits), Expect = 3.2e-15, Sum P(2) = 3.2e-15  
Identities = 15/57 (26%), Positives = 19/57 (33%)

Query: 22 TVHEPVVTVQAVHPPAPAHPSLLDKMEKAPPOQHEG-LKSKEHLPOQPAEGKTASR 77  
T EP T P P P+ E AP P+ +KE P T +  
Sbjct: 377 TPKEPAPTTKSAPTTKPEPAPTTKPEPAPTTKPEPAPTTTKEPAPTTTKSAPTTK 433

Score = 46 (6.9 bits), Expect = 4.0e-15, Sum P(2) = 4.0e-15  
Identities = 16/58 (27%), Positives = 22/58 (37%)

Query: 20 LATHEPVVT---QWAVHPPAPAHPSLLDKMEKAPPOQHEGLKSKEHLPOQPAEGKT 74  
L T EP T + A P P+ + P +P KS P++A T  
Sbjct: 344 LTTPEPTPTTKEPASTTKPEPTTTIKSAPTTKPEPAPTTTKSAPTTTKEPAPTTT 401

Score = 42 (6.3 bits), Expect = 1.0e-14, Sum P(2) = 1.0e-14  
Identities = 15/60 (25%), Positives = 21/60 (35%)

Query: 22 TVHEPVVTVQAVHPPAPAHPSLLDKMEKAPPOQHEGLKS-KEHLPOQPAEGKTASRRVP 80  
T EP T P P P+ + AP P+ + KE P E + + P  
Sbjct: 463 TPKEPAPTTKEPAPTTKPEPAPTAPKKPAPTTKPEPAPTTKPEPAPTTTKEPSPTTKEP 522

Score = 39 (5.9 bits), Expect = 2.1e-14, Sum P(2) = 2.1e-14  
Identities = 15/55 (27%), Positives = 20/55 (36%)

Query: 22 TVHEPVVTVQAVHPPAPAHPSLLDKMEKAPPOQHEGLKSKEHLPOQPAEGKTAS 76  
T EP T P P+ + P +P KS ++PA T S  
Sbjct: 494 TPKEPAPTT----PKEPAPTTKPSPTTKEPAPTTTRESAPTTTKEPAPTTTKS 544

Pedant information for DKF2phtes3\_4ol9, frame 2

Report for DKF2phtes3\_4ol9.2

[LENGTH]	1180
[MW]	127693.40
[PI]	10.25
[HOMOL]	SWISSPROT:MUC2_HUMAN MUCIN 2 PRECURSOR (INTESTINAL MUCIN 2). 1e-08
[FUNCAT]	98 classification not yet clear-cut [S. cerevisiae, YJR151c] 6e-06
[FUNCAT]	30.01 organization of cell wall [S. cerevisiae, YI019c] 6e-06
[FUNCAT]	30.90 extracellular/secretion proteins [S. cerevisiae, YI019c] 6e-06
[FUNCAT]	01.05.01 carbohydrate utilisation [S. cerevisiae, YI019c] 6e-06
[BLOCKS]	BLO0412B Neuromodulin (GAP-43) proteins
[PROSITE]	CYTCHROME_C 1
[PROSITE]	MYRISTYL 12
[PROSITE]	CAMP_PHOSPHO_SITE 1
[PROSITE]	CK2_PHOSPHO_SITE 8
[PROSITE]	PKC_PHOSPHO_SITE 25
[PROSITE]	ASN_GLYCOSYLATION 2
[RM]	Alpha Beta
[RM]	LOW_COMPLEXITY 5.00 %

SEQ MTLOGRADLSGNQNAAGRLATVHEPVVTQMAVHPAPAHPSLLDKMEKAPQPQHEGLK  
SEG .....  
PRD ccc  
SEQ SKEHLPOQPAEGKTASRRVPLRAVVESQAFKNI LVDENDMHARAATLIQANRWGYWLR  
SEG .....  
PRD ccc  
SEQ QKLISQMAKAKIQAMRRFNKRHLNHSKSLVKKTRAEEDIPYHAPQVRFQHPENR  
SEG .....  
PRD hhh  
SEQ LLSPPIMWNKETQFPCDNLVLCRQSSPLLQPPAAQGTPEPCVQCPHARVRGLAFLPH  
SEG .....  
PRD ecc  
SEQ QTVTIRFPCPVSLDAKQPCLLTTRTIRSTCLVHIEGDSVKTARVARTHKARAPETPLSR  
SEG .....  
PRD ecc  
SEQ RYDQAVTAPBRAQTQGPVKAETPAFQICPGPHITKTLQTVPVVSVTLPTYPASTHT  
SEG .....  
PRD ccc  
SEQ TTPKTSPPVKYTIKTFAQMYPGPTVTKTAPHTCPMPTMTKI QVHPTASRTGTPTOTCP  
SEG .....  
PRD xxxxxxxxxxxxxxxx  
SEQ ATITAKNRPPQVSLASINKSLFQVCPGPANAKTPPQHWVPTPARKPLQTLCSATNSKTS  
SEG .....  
PRD ccc  
SEQ SQRSFVGVTKPSQTRL PANITKTFAQLRSVATILKTLCLASPTVANVKAQPPVAVAGT  
SEG .....  
PRD ccc  
SEQ PNTSGSIHENPPKAKATVWVQAARVVRASSPSYLAEGKIRCLAPHPTQVPRAAELP  
SEG .....  
PRD ccc  
SEQ LEAEKIRGTGQAKTMAFKTSVAVEMAGAPSWTKVAEGDKPPHYVVPVDMVTLPRG  
SEG .....  
PRD ccc  
SEQ QLAAPLTNASSQRHPPCLSORPLAAPLTKASSQGHLPTELTKTPSLAHLDTCLSKMSQT  
SEG .....  
PRD ccc  
SEQ HLATGAVKVSQAPLATCLTKTQSRGQPTIDITTCCLI PAHQAADLSSNTHSQVLLTGSYV  
SEG .....  
PRD ccc  
SEQ SNHACQRLGLSAPPMAKPEDRQTQOPHGHVPGKTTQGGPCPAACEVQGHVPPMAPTG  
SEG .....  
PRD ccc  
SEQ HSTCHVESHGNGATRAQPSHFGQAVPCQEDTGADAGVVGQSMWRAMEPARGAASVOT  
SEG .....  
PRD ccc  
SEQ WRNKAVVPPRRSGEPHYSMAAEEIRI LAVITIQAGVGYLARRRIRLMWRGAMVQATW  
SEG .....  
PRD ccc  
SEQ RGYRVRRNLALHLCRATTTIQSAMRGYSTRDQARHWQNLHPVTMVELGSRAGVMSDRSWF  
SEG .....  
PRD hhh  
SEQ QDGRARTVSDHRCFQSCQAHACSVCHLSLRIGSPSPVVMVGSSPRTCHTCGRTOPTRV  
SEG .....  
PRD hcc  
SEQ VQGMGQTEGPGAVSWASAYQLAALSPPQPHRQDKAATAIQSAMRGFKIROOMRQGMMA  
SEG .....  
PRD ecc  
SEQ KIVQATWRGHHTRSCLKNTALGFPADPSASSRHOWPGI  
SEG .....  
PRD hhh

## Prosite for DKFZphtes3\_4ol9.2

PS00001	542->546	ASN_GLYCOSYLATION	PD0C00001
PS00001	668->672	ASN_GLYCOSYLATION	PD0C00001
PS00004	282->286	CAMP_PHOSPHO_SITE	PD0C00004
PS00005	76->79	PKC_PHOSPHO_SITE	PD0C00005
PS00005	148->151	PKC_PHOSPHO_SITE	PD0C00005
PS00005	244->247	PKC_PHOSPHO_SITE	PD0C00005
PS00005	265->268	PKC_PHOSPHO_SITE	PD0C00005
PS00005	278->281	PKC_PHOSPHO_SITE	PD0C00005
PS00005	281->284	PKC_PHOSPHO_SITE	PD0C00005
PS00005	285->288	PKC_PHOSPHO_SITE	PD0C00005
PS00005	288->291	PKC_PHOSPHO_SITE	PD0C00005
PS00005	299->302	PKC_PHOSPHO_SITE	PD0C00005
PS00005	322->325	PKC_PHOSPHO_SITE	PD0C00005
PS00005	414->417	PKC_PHOSPHO_SITE	PD0C00005
PS00005	424->427	PKC_PHOSPHO_SITE	PD0C00005
PS00005	481->484	PKC_PHOSPHO_SITE	PD0C00005
PS00005	610->613	PKC_PHOSPHO_SITE	PD0C00005
PS00005	671->674	PKC_PHOSPHO_SITE	PD0C00005
PS00005	679->682	PKC_PHOSPHO_SITE	PD0C00005
PS00005	900->903	PKC_PHOSPHO_SITE	PD0C00005
PS00005	959->962	PKC_PHOSPHO_SITE	PD0C00005
PS00005	987->990	PKC_PHOSPHO_SITE	PD0C00005
PS00005	1015->1018	PKC_PHOSPHO_SITE	PD0C00005
PS00005	1049->1052	PKC_PHOSPHO_SITE	PD0C00005
PS00005	1065->1068	PKC_PHOSPHO_SITE	PD0C00005
PS00005	1106->1109	PKC_PHOSPHO_SITE	PD0C00005
PS00005	1146->1149	PKC_PHOSPHO_SITE	PD0C00005
PS00005	1171->1174	PKC_PHOSPHO_SITE	PD0C00005
PS00006	22->26	CK2_PHOSPHO_SITE	PD0C00006
PS00006	42->46	CK2_PHOSPHO_SITE	PD0C00006
PS00006	156->160	CK2_PHOSPHO_SITE	PD0C00006
PS00006	546->550	CK2_PHOSPHO_SITE	PD0C00006
PS00006	848->852	CK2_PHOSPHO_SITE	PD0C00006
PS00006	988->992	CK2_PHOSPHO_SITE	PD0C00006
PS00006	1003->1007	CK2_PHOSPHO_SITE	PD0C00006
PS00006	1027->1031	CK2_PHOSPHO_SITE	PD0C00006
PS00008	11->17	MYRISTYL	PD0C00008
PS00008	14->20	MYRISTYL	PD0C00008
PS00008	539->545	MYRISTYL	PD0C00008
PS00008	591->597	MYRISTYL	PD0C00008
PS00008	746->752	MYRISTYL	PD0C00008
PS00008	777->783	MYRISTYL	PD0C00008
PS00008	853->859	MYRISTYL	PD0C00008
PS00008	878->884	MYRISTYL	PD0C00008
PS00008	882->888	MYRISTYL	PD0C00008
PS00008	1008->1014	MYRISTYL	PD0C00008
PS00008	1053->1059	MYRISTYL	PD0C00008
PS00008	1083->1089	MYRISTYL	PD0C00008
PS00190	1042->1048	CYTTOCHROME_C	PD0C00169

(No Pfam data available for DKFZphtes3\_4ol9.2)



WO 01/12659

PCT/IB00/01496

DKF2phtes3\_50j4

group: testes derived

DKF2phtes3\_50j4 encodes a novel 187 amino acid protein proline rich protein.

No informative BLAST results; No predictive prosite, pfam or SCOP motifs.

The new protein can find application in studying the expression profile of testis-specific genes.

unknown, prolin rich protein

complete cDNA, complete cds, EST hits

Sequenced by DKF2

Locus: unknown

Insert length: 1186 bp

Poly A stretch at pos. 1176, polyadenylation signal at pos. 1126

```
1  CACTGGGGCT CTGAAGCTCA GAGCTCACCC CTGAGATGGG CTCTCTTAGG
51  CCTCTCTGGA TGAGGGAGCC ACCAGGACCC AGTCTGTGA TGCTGCTCT
101 TCCCTCTACC AGCAGCTGCC GCGCCAGAGA CCAGGGCACC CCTGAAGTCC
151 AGCCGACGCC TGCAGAGGAC ACATGGAGGG GCAGGGGGCC TGGATCCGAG
201 CAGGAGAACCC CAGAGAGCCA GCCTCAGAAG AGGCCAGGCC CCTCAGCCAA
251 GCGCTCCCTC CTAGCTGAGG TCAAGGGCAG CGTCTGGGCC AGCGAACAGG
301 GCACCTTGAA TCCACGGGCT CAAGACCCCT TCCAGCTCTC CGCTCTGGG
351 GTCTCTCTTA AGAGGGCTGC AATGTGTGTG CTCAGTGGCC TCACCCCTTT
401 CTACAAGGAG GGCAAGTTTG CTTCAGAGGA GTTGTATAA GGCTTTGGCC
451 GGCACCTCTC ACACCTTGCTG ACTCAGAAGA CCTCTCTGGG AAGGAGCGTG
501 AAGAAGAGAG CCCAGAACCT CATCAGGCAC TTCTTCCATG GCGGGGCCCG
551 GTCCGAGAGC GAAGCTGACT GGCATGGCTT GTGTGGCCCC CAGAGATGAC
601 CACTGCTGGG CTGGGAGGGG CCGGCTCTCT CCCCAGATT CTAGCATGGG
651 TCATCTCTGG CCTCACCTGC TGATGCCAGG GGCATGCTT TTTCTAGTC
701 TTCTCTCTTT CCAACCATAC TTGGCTTTGG GGATGACCCG AGACACCCCC
751 TCAATCCAGG TCAAGAGTCA GCGCACCTTT TTCTCTGCTT GCAGAGCTTA
801 TAGACCTTTC TCAGAGCGGT CCTCATGGCT GGCTTTCTG GGACACATGT
851 CGAGGACAGA AGGTGGAGGG TGGTGGAGCT GCTGCTGGAA GAAGGGGAAG
901 GAAGAGTGGC CCTCCGCCGA GTTCTAAGTC AGGATGAGGC CCACCTGTCC
951 AAGGTATCGG AACTACCCA GGGGAGCCCT AGATCTCCGA CCCACTGCC
1001 CATCTATTAC GATGCGAGCT TCCAGCCTTG CCGAGTCAAG AGCTGTGGCA
1051 GAGGAGAGGC AGCCAGGCCG TGTCTCTGCT CAGCTCTCTG TCAGGAAGGC
1101 CAGGCGCTGAC AGATGTTTGG GAGAGGAATA AAGTTGTGTT GTTGTGGGGC
1151 ATGACGGCGT GCACACAGCC CTTTCAAAA AAAAAA
```

BLAST Results

No BLAST result

Medline entries

No Medline entry

Peptide information for frame 3

ORF from 36 bp to 596 bp; peptide length: 187  
Category: putative protein

```
1  MGSPRPQMR EPPGPSAVMP ALPSTSTCPP RDQGTPEVOP TPAKDTWKGK
51  RFRSQENPE SYPKRRPFP AKFSVVAEVK GSVSESDST LMPFAGOPTQ
101  LSAPCVLKE ANPVVVKLT PYKEGRFAS KEIFRGFARN LSKILTQRTS
151  PGRSVKEAQ NLIRHFFHGR ARCESEADWH GLCGPQR
```

BLASTP hits

Entry MMU92455\_1 from database TREMBL:

WO 01/12659

PCT/IB00/01496

product: "WW domain binding protein 7"; Mus musculus WW domain binding protein 7 mRNA, partial cds.  
Score = 134, P = 6.9e-08, identities = 45/125, positives = 56/125

Alert BLASTP hits for DKFZphtes3\_50j4, frame 3

No Alert BLASTP hits found

Pedant information for DKFZphtes3\_50j4, frame 3

Report for DKFZphtes3\_50j4.3

[LENGTH] 187  
[MW] 20353.06  
[pI] 9.76  
[PROSITE] MYRISTYL 1  
[PROSITE] AMIDATION 1  
[PROSITE] CK2\_PHOSPHO\_SITE 6  
[PROSITE] PKC\_PHOSPHO\_SITE 6  
[KW] All\_Alpha  
[KW] LOW\_COMPLEXITY 8.56 %

SEQ MGSPRPFGHREPPGFSVMPALPSTSTCPRDQGTPEVQPTPAKDTWKGRPRSQENPE  
SEG xxx  
PRD ccc  
SEQ SQPQRPRPSAKPSVVAEVKGSVSASEQGLNPTAQDPFOLSAPGVSLKEANVVKCLT  
SEG .....  
PRD cccccccccccccchhhhhccccccccccccccccccccccccccccchhhhhheeecc  
SEQ PFYKEGKFASKELFKGFARHLSKLLTQKTSFGASVKEEAQNLIRHFTGPARCEADWH  
SEG .....  
PRD cccccchhhhhhhhhhhhhhhheccccchhhhhhhhhhhhhccchhhhhhhhh  
SEQ GLCGPQR  
SEG .....  
PRD ccccccc

Prosite for DKFZphtes3\_50j4.3

PS00005	3->6	PKC_PHOSPHO_SITE	PDOC00005
PS00005	46->49	PKC_PHOSPHO_SITE	PDOC00005
PS00005	70->73	PKC_PHOSPHO_SITE	PDOC00005
PS00005	107->110	PKC_PHOSPHO_SITE	PDOC00005
PS00005	146->149	PKC_PHOSPHO_SITE	PDOC00005
PS00005	154->157	PKC_PHOSPHO_SITE	PDOC00005
PS00006	54->58	CK2_PHOSPHO_SITE	PDOC00006
PS00006	84->88	CK2_PHOSPHO_SITE	PDOC00006
PS00006	94->98	CK2_PHOSPHO_SITE	PDOC00006
PS00006	107->111	CK2_PHOSPHO_SITE	PDOC00006
PS00006	134->138	CK2_PHOSPHO_SITE	PDOC00006
PS00006	175->179	CK2_PHOSPHO_SITE	PDOC00006
PS00008	81->87	MYRISTYL	PDOC00008
PS00009	48->52	AMIDATION	PDOC00009

(No Pfam data available for DKFZphtes3\_50j4.3)

WO 01/12659

PCT/IB00/01496

DKFZphtes3\_50n06  
-----

group: testes derived

DKFZphtes3\_50n06 encodes a novel 186 amino acid protein without similarity to known proteins.

No informative BLAST results: No predictive prosite, pfam or SCOP motif.

The new protein can find application in studying the expression profile of testis-specific genes.

unknown

complete cDNA, complete cds, EST hits

Sequenced by DKFZ

Locus: unknown

Insert length: 1095 bp

Poly A stretch at pos. 1085, polyadenylation signal at pos. 1061

```
1  CAAGACCCCTC  GGAGCCAAGA  AACAACTCTG  AGTTCCAGAT  TTCGGAAGGT
51  TCACGAGTGT  TGCCGACACG  CCTCCCAAC  TGCAGACATC  CTCCTTGGAG
101  GACCTGCTGT  GCTCACAATG  CCCCCTGTCC  AGCAGAGACG  ACACCTCCCC
151  GGGCTGTGCA  GCCCCCTCCC  AGCAACCTTC  CAAGGCTTTC  CTCAGTCCCC
201  CAGAGCCACA  TAGCCACCGA  GGCACCGACA  GGAAGCTGTC  CGCCCTCCTG
251  AGCCCCCTTC  AAGACTCACT  GGTGGACAAG  ACCCTGCTGG  AGCCCAAGGA
301  GATGGTCCGG  CCTAAGAAGG  TGTGTTTCTC  GGAGAGCAGC  CTGCCACCCG
351  GGGACAGGAC  CAGAGAGAGC  TACTACCTCA  ATAGATCCCA  GAGCTTCGGG
401  GCGCCGAGAA  AGGACGCGCG  CGTGGTGGGC  GAGATCGCCT  TCCAGCTGGA
451  CCGCCGCGATC  CTGGCCTACG  TGTTCCTGGG  CGTGACGCGG  CTCTACGGCT
501  TCACGGTGGC  CAACATCCCC  GAGAAGATCG  AGCAGACCTC  CACCAAGTCT
551  CTGGAGCGCT  CGCTGGACGA  GAGGAAGCTG  CGCAGACTCA  CGCAGCGCTA
601  CTTGGCTCTG  AGCCGCGCGC  TGGAGAAGCT  CGGCTACAGC  CGCGACGTGC
651  ACCCGGGGTT  CAGCGAGTTC  CTCATCAACA  CCTACGGAAT  CCTGAAGCAG
701  CGGCCCGACC  TGCGCGCCAA  CCCCCTGCAC  AGCAGCCCGG  CGCGCGTGGG
751  CAAGCTGGTC  ATCGAGCTGG  TGCCCCCA  GTTCTGGGG  GACTCGCTGC
801  TCTGTCTGAA  CTCCTGTGTC  GAGCTTCCCA  AGGAGGACGG  CAAGCCCTCT
851  TTCGCTGTGT  GAGCCGCCCC  GCGCCCGCGC  CTTTGCCTGC  AGTAAACCGG
901  TTTGTTCCAA  CCGCGGCGCG  CGGTGCTCTC  TCGCGCTCCC  CCGGAGGGGG
951  AAAAGGGCGC  GTCCCCCGCG  CCGAGGCCCA  GAGAAAGCCC  CGCTCCACCC
1001  GGTGCTGGGC  CCGACCGCCA  GCCCCCGGCT  GCGCGACCTC  GCGGAGTGCT
1051  TCTCACCCCT  CATTAAATC  ATCGGTTTGC  TTGTCAAAAA  AAAAA
```

BLAST Results  
-----

No BLAST result

Medline entries  
-----

No Medline entry

Peptide information for frame 2  
-----

ORF from 302 bp to 859 bp: peptide length: 186  
Category: putative protein  
Classification: no clue

```
1  MVRPKKVCFS  ESSLPDGRD  RRSYYLNEIQ  SFAGAEDAR  VVGEIAFOLD
51  RRILATVFPQ  VTRLYGPTVA  NIPEKIEQTS  TNSLDGSVDE  RKLRELTORY
101  LALSARLEKL  GTSRDVHPAF  SEFLINVTGI  LKQRPDLRAN  PLHSSPAALR
151  KLVIDVVPFK  FLGDSILLLN  CLCELSREKG  KPLFAV
```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphtes3\_50n06, frame 2

No Alert BLASTP hits found

Pedant information for DKFZphtes3\_50n06, frame 2

Report for DKFZphtes3\_50n06.2

```
[LENGTH]      186
[MM]           21049.39
[PI]           9.28
[KW]           All_Alpha
[KW]           LOW_COMPLEXITY      5.38 %

SEQ  MVRPKKVCFSSESLPTGDRTRRSYVLNEIQSFAGAEDARVVGEIAFQDORRILAYVFP
SEG  .....
PRD  ccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc

SEQ  VTRLYGFTVANIPKIEQTSTKSLDGSVDERKRLRELQRYLALSARLEKLGYSRDVHPAF
SEG  .....
PRD  ccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc

SEQ  SEFLINTYGIKQRPDLRANPLHSSPAALRLVIDVPPKFLGDSLLINCLCELSKEDG
SEG  .....
PRD  hhhhhcccccccccccccccccccccccccccccccccccccccccccccccccccccc

SEQ  KPLFAN
SEG  .....
PRD  ccccc
```

(No Prosite data available for DKFZphtes3\_50n06.2)

(No Pfam data available for DKFZphtes3\_50n06.2)

WO 01/12659

PCT/IB00/01496

DKF2phtes3\_50n23

group: testes derived

DKF2phtes3\_50n23 encodes a novel 499 amino acid protein without similarity to known proteins.

No informative BLAST results; No predictive prosite, pfam or SCOP motif.

The new protein can find application in studying the expression profile of testis-specific genes.

unknown

2 EST hits  
(from other testis libraries) testis specific cDNA?

Sequenced by DKF2

Locus: unknown

Insert length: 1907 bp

Poly A stretch at pos. 1897. polyadenylation signal at pos. 1872

```
1 GGGCACCAGC CACTTTCAC CATGACTGTG CGCTCGAGGG TCGCAGATGT
51 GTTCGGCAGC AAGGACACTG AGAGCCTTGA GCCTGTGCTT TTACCCCTTA
101 TAGATCGCAG GTTTCCTAAG AATGGGAAA GACCGGTGGC AGAAAGCTTA
151 GGGCACAAG ACAAGACCA GGAGGACTAC TTCCAGAGGG GAGGACTCCA
201 AATTAGTTC CACTGTAGCA AGCAGCTGTC TCTAGAGGC TCAGCGAGG
251 TGACCTCTGA GAGCAAGAG CAGCCTTGGG AGGAGGAATT CGGCCGGGAG
301 ATCGGAGGCG AGCTGTGGCT GGAGGAGGAG GAGATGTGGC AGCAGCGGCA
351 GAAGAAGTGG GCCCTGCTGG AGCAGGAGCA TCAGGAGAAG CTGCGGCAGT
401 GGAATCTGGA AGACCTGGCC AGGAGAGAAC AGCGAGATG GGTCCAGCTA
451 GAAGAAGAGC AGGAGAGGCC AGCGAGAGAG CCAGAGCAGC TAGGGGAGGA
501 TGTGAGAGGG AGGATCTTCA CACCCACCAG TCGATGGAGG GACTTGGAGA
551 AGCGAGAGCT ATCATTAGTG CTTGCCCCAA GCCGAGCCCA ATCTGCTCAC
601 CAAAGCAGGA GGCCACACTT GCCCATGTCT CCTAGTACCC AGCAGCCTGC
651 CTTGGGAAG CAGAGACTTA TGAGTTCAGT GGACTTTACG TACAGACCCG
701 GGACCCGCCG AGTTCGCCCA AAGCCCAAGA AATCTGCTCT CTTTCTGTCT
751 ACTGGGACAT CCATCCGAGG GCTGACCTGG CCCTCTTTGC AGATATCCCC
801 TGCAATATT AAGAAGAGG GTTACCACAT GGACATGGAG GCCCAGAGGA
851 AGAAGCTGGA CTCTCTGAGT GAGGAGCTGT AGTTGAGGCT GCCCAGTAC
901 CTGCGAGGCA AAGCACTGGA GCTCACCACC ACCACATGG AGCTGGGCGC
951 GCTCAGGCTG CAGTACCTGT GCCATAAGTA CATCTTCTAT AGACGCTCTC
1001 AGAGCTCTCG GCAAGAAGCG ATCAACCATG TACAAATCAT GAAGAAGAG
1051 GAGGCTTCTT ACAAGGCCCA GAGCTCTTAT ATCTCTCTGG AAACATTGTA
1101 CCGCTCGCAG AGTCTCAGGC TGCAGGCTCG GACGGACAAG CAGAAGGGGC
1151 TGGAGGAGAA GCACCGAGAG TGCTTGAGCA GCATGGTGAC CATGTTCCCC
1201 AAGCTCCAGC TGGAGTGGAA GCTTCACCTG AACATCCTGT AGGTCACTCT
1251 GCGAAGCCA AAGAAATGCA AGTGTGCTGG AGCTCGGCC CGGCAATCTC
1301 GCCCAGTGG CCCCACCTAC AAGCAGCCCT TTCTGTCTAG GCACCGGCA
1351 TGTGTGCCCC TGCAGATGGC CCGCCACACG GGAAGCAGA TGGAGGCTGT
1401 CTGGAAGACC GAGGTGGCCT CTCTCAGTTA CGCAATAGAA AAAAGACCCC
1451 CTGCGAGCCT TCCCGGGAC CAGCTGAGGG GAGACGAGA TATTCCCGGG
1501 CTCTTGACAC TGGAGCTGTA GTCTCTCTCG CACAAAGCCG TGAATCTCTT
1551 GAAGGCCAGC TAAGCGCCTC AGCGAACCAA AGGAAGGAAT GCCAGGAACC
1601 TACAAATGAA TCCGCTTAGC TTGTTCAAAA AAAGTCAAGG GAGTCACTCC
1651 CTGGAAGCCA AATFAGCCAG AAGGATCAAG ACAGCCCGAG TCTCCACTGC
1701 ATCCCTCAGC CAGTGATTCT CAACCTTCTG AGGGACGAAA ACCCAGAGAG
1751 AACTTGGTCA AATGACAGGT TCCCGAGTGG TGCTTTTAAA GAAACCTCTT
1801 GGGGGTGTCT GAGTACTCCT AGAATTTTGA GAAACACTGC TTCCCTCTCTG
1851 CAGTCCCAA ACTCTACATT TTAATAAAT AGAGGTGTGT TTATTTTAAA
1901 AAAAAAA
```

#### BLAST Results

No BLAST result

#### Medline entries

No Medline entry

## Peptide information for frame 1

ORF from 22 bp to 1518 bp; peptide length: 499  
Category: similarity to known protein  
Classification: no clue

```
1  MTVRSRVADV FGSKOTESLE PVLLPLVDOR FPKKWERPVA ESLGKKDKDQ
51  EDYFQGGGLQ IKFHCSKOLS LESSRQVTSE SODEPWEEEF GREHRRQLML
101 EEEEMWQOQK KMWALLEQEH QKLRQWLE DLAREQORRW VOLEKEQESP
151 RREPEQLGED VERRIFTPTS RWRDLEKAEI SLVPAPSRTO SAHOSRRPHL
201 PMSPTQOPA LGKORPNSSV EFTYRPRTRR VTPKPKSAS FPVTGTSIRR
251 LTMPSLOISP AMIKKVVYHN DNEAQRMNQ LLSSESLRL PHYLRKALE
301 LTTTWELGA LRQLVCHNY IFYRRLQSLR QEAINHVQIM KETASVHAQ
351 NLVIFLENIQ RLQSLQAN TDKQGLEEK HRECLSMVT MFPKLQLENN
401 VHLNIPVETS PKPKCKLPA ASPKHIFPSG PTYKOPFLSR HRACVPLQNA
451 RQQKQMEAV WKTEVASSY ATEKTPASL PRDQLRGHPD IPRLLTLQV
```

## BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZp385t030n23, frame 1

PIR:S28589 trichohyalin - rabbit, N = 1, Score = 134, P = 5.3e-05

TREMBL:AF132479.1 product: "Ese2L protein"; Mus musculus Ese2L  
protein mRNA, complete cds., N = 1, Score = 130, P = 0.00017

>PIR:S28589 trichohyalin - rabbit  
Length = 1,407

## HSPs:

Score = 134 (20.1 bits), Expect = 5.3e-05, P = 5.3e-05  
Identities = 88/354 (24%), Positives = 154/354 (43%)

```
Query: 29 RRFPKKWERPVAESLGHKKDKDQEDYFQGGGLQIKFHCSKOLSLESSRQVTSESODEPWEE 87
      R++ K +R + L + ++E ++ G + F +OL +++ E +EE +
Sbjct: 165 RQYRDKEQLRQQLKEERRAAEEQLRRRGRDAEEFTEEEQLRRRQQLKRELREEEQ 224

Query: 88 EEFGRMRQLMLEEEEMWQOQKMWALLEQEHQKLRQWLEDLAREQORRWVOLEKEQ 147
      RE + L+EEE RQ++W E Q++LR+ LE++ RE+++R Q E+ +
Sbjct: 225 RREERQHERA-LQEEECOLLQRRRWRE-EPREQQQLRR-ELEET-REREQLEQEERRE 280

Query: 148 ESPPREPEQLGEDVERRIFTPTSRWDLKAEISLVAPASRTQSAHOSRRPHLPMSPTQ 207
      + RRE ++L E ERR ++ + E L R O Q R + +
Sbjct: 281 QQLRRE-RL-QDEERREQLRRELEEI REREQLEQEERREQLEQEERREQLKRELE 338

Query: 208 QPALKGKORPNSSVEFTYRPRTRVTPKPKSASFPVTGTSIRRLTMPSLOISPANIKK-K 266
      + +QR +E R R + + + +A G S+ R W SA ++ K
Sbjct: 339 EIREREQ---LEQEER-REQLAEVREQAR--ERGSLTR-RWQOLESEAGARQSK 390

Query: 267 VYHMDAQQRNQLLSESELRLPHYLRKALELTTTNN-----ELGALRLQYLCHRY 320
      VY +R+ Q L ++ E R R + LE E R Q L +
Sbjct: 391 VYS---RPRRQEQSLRQDQERR-ORQERELEEDQARRQQQQAEESEARRRQLSARP 446

Query: 321 IFYRRLQSLRQEAINHVOIMKETASVKAQNLVI-FLENIQLQSL-RLQMTDKQKGLE 378
      R Q +E Q +E E + + + FLE ++LQ R Q ++ E
Sbjct: 447 SLRER-QLRAGEERQEQQRFREEEQRRERQELQFLKEEQQLRRERAAQQLQEDSFOE 505

Query: 379 EKHR 382
      ++ R
Sbjct: 506 DRER 509

Score = 119 (17.9 bits), Expect = 2.2e-03, P = 2.2e-03
Identities = 79/357 (22%), Positives = 150/357 (42%)

Query: 33 KKWERPVAESLGHKKDKDQEDYFQGGGLQIKFHCSKOLSLESSRQVTSESODEPWEEFGR 92
      ++ E+ + + E ++E Q+ + + + Q R+ + + EE+ + +
Sbjct: 990 RREQLRQERDRKFREEQLQGE---REEELRRQERDRKFREEERQLRQLEEQFRQ 1046

Query: 93 ENRRQLMLEEEEMWQOQKMWALLEQEHQKLRQWLEDLAREQORRWVOLEKEQESPRR 152
      E R+ LEE+ + Q+++K L QE K R+ E+ R +Q R QL +E+ + R
Sbjct: 1047 ERDRKFRLEQ-IRQEAQK-QLRRQERDRKFRE---EQQRFRQEREQLRERDRKFR 1101

Query: 153 EPEQLGEDVERRIFTPTSRWDLKAEISLVAPASRTQSAHOSR--RPHLPMSPTQOPA 210
```

Query: 1102 E EQL ++ E R R L + E L + + + R R + + +  
Sbjct: 1102 EEEQLQEEERLRQERARKLREE-OLLRREEQLLRQERDRKFREEQLQEESEER 1160

Query: 211 LGRQ--RPHSVFTYRPRTRVPTKPKXSASFVPTGTSIRLWPSLQISPAKIKKV 267  
Sbjct: 1161 LRRQERERKLREEQLQEEERLRQERARKLREEQLRQEEQLRQERARKLREE 1220

Query: 268 YHMMEQ-----RKLQLLS-ESELRPLPHYLASKALELTTTTLGALRLQYL 316  
Sbjct: 1221 QLLRQEEQLRQERDRKFREEQLLRQEEQLLRQERDRKFREEQLLQEEERLRQER 1280

Query: 317 CHKYIFYRKLQSLRQEAINHVOIKHTEASYNAQNYIFLENDRLQ-SIALQANTOKK 375  
Sbjct: 1281 ARK--LREEEQQLFEEQERLQERDRYRAEQFAREKSRRLERLQEEEQRRR 1338

Query: 376 GLEKHRE 383  
Sbjct: 1339 ERERKPRE 1346

Score = 109 (16.4 bits), Expect = 1.9e-01, P = 1.7e-01  
Identities = 37/113 (32%), Positives = 60/113 (53%)

Query: 67 KQLSESSROVTSESO--EEPWEETGEMRRQLMLEEENMOOROKKWLLEQEHQKL 124  
Sbjct: 764 QQLAREDRKFREEQLQEEERLRQERARKLREEQLQEEES-ALRQERERKL 822

Query: 125 ROWNLEDLAREQQRWVQLEKEQSPREPEOLGEDVERRITPTSRWRLEKAE 179  
Sbjct: 823 REE--EQLQEEERLR-RQERERKLREEQLRQEEQLRQEEERLRQERERKLREE 872

Score = 107 (16.1 bits), Expect = 3.0e-01, P = 2.6e-01  
Identities = 35/109 (32%), Positives = 61/109 (55%)

Query: 71 LESSROVTSSEQEPWE-EETGEMRRQL--WLEEEEMOOROKKWLLEQEHQKL 126  
Sbjct: 742 LREEQLQEESEERLRQERERKLREEQLRQERDRKFREEQLQEEER-RLRQERERKLRE 800

Query: 127 WNLLEDLAREQQRWVQLEKEQSPREPEOLGEDVERRITPTSRWRLEKAE 179  
Sbjct: 801 E--EQLQEEERLR-RQERERKLREEQLQEEERLRQERERKLREE 850

Score = 104 (15.6 bits), Expect = 9.4e-02, P = 9.0e-02  
Identities = 84/339 (24%), Positives = 149/339 (43%)

Query: 67 KQLSESSROVTSSEQEPWEETGEMRRQL-WLEEEEMOOROKKWLLEQEHQKL 123  
Sbjct: 451 RQLRAEERQEQRFREE--EQRRERQEQLOFLEEEQLQRRERAQQLQEDSFQEDR 507

Query: 124 LACNLEDLAREQQRWVQLEKEQSPR---EP--EQLGEQVE-RRITPTSRWRDL 175  
Sbjct: 508 ERNRQEQQRPGQTRW-QLQEEAQRARHTLYAKPGQEQLEEEELQREKRRQEREREY 566

Query: 176 EKAEKSLVPASRTSAGSRAHPLWSPSTQOPALGKORPMSSVFTYRPT---RRV 231  
Sbjct: 567 REE-KLQREDEKRRQERERQYRELEELQEEQL-NDRKLREEQLQEEERLRQ 624

Query: 232 PTKR--KSASFVPTGTSIRALTWPSLQISPAKIKKVYHMMEQK--NLQLSEE 285  
Sbjct: 625 ERERKLREEQLRQEEQLRQERERKLREEQLLRQEEQLRQERERKLREEQLQER 684

Query: 286 SELALPHYLRSKALE-----LTTTTLGALRLQYLCHKYIFYRRL-QSLRQEAINH-- 337  
Sbjct: 685 EERLRQERARKLREEQLRQEEQLRQERERKLREEQLLRQEEQLRQERDRKLRE 744

Query: 338 --QMKHTEASYNAQNYIFLENDRLQSLRQANTOKKGLEKHRECL 385  
Sbjct: 745 EQLQEESEERLRQ-----EREQLRERDRKFREEQLQEEERL 789

Score = 103 (15.5 bits), Expect = 1.2e-01, P = 1.1e-01  
Identities = 42/152 (27%), Positives = 74/152 (48%)

Query: 36 ERPAESLGHKDKQEDYFKQGLQIKFHCSKQLSESSROVTSSEQEPWEETG-REM 94  
Sbjct: 835 ERLRQERERKLREEQLRQEEQLRQERARKLRE-EEQLRQEEQLRQERDRKLEE 893

Query: 95 RRLMLEEEMMOOROKKWA-----LLEQEHQKLRWNLEDLAREQ--RRVQ-LEKE 146  
Sbjct: 894 EQLRQEEQLRQERDRKLEEQLQEESEERLRQERERKLREEQLLRQEEQLARE 953

Query: 147 QESPRPEPEOLGEDVERRITPTSRWRLEKAE 179  
Sbjct: + RE EQL ++ E R R L + E

Sbjct: 954 RARKLREEEQQLQEREERLRQERARKLREE 986  
Score = 103 (15.5 bits), Expect = 7.8e-01, P = 5.4e-01  
Identities = 31/91 (34%), Positives = 52/91 (57%)  
Query: 67 KQLSLESSQVTSQEEFWEFEFGREHRRQLWLEEEEMWQORQKXWALLEQEHQEKLRQ 126  
++L E R++ E Q EE+ R+ R + EEE++ Q+R+++ L QE KLR+  
Sbjct: 642 QELRQERERKLREEEQQLRREEQLRQERERKLREEEQQLQEREER-LRQERARKLREE 700  
Query: 127 WNLEDLAREQQRWVLEKEQESPRREPEQL 157  
E L R+++ +L +E+E RE EQL  
Sbjct: 701 E--EQLLRQEEQ---ELRQERERKLREEQL 726  
Score = 101 (15.2 bits), Expect = 2.0e-01, P = 1.8e-01  
Identities = 38/111 (34%), Positives = 57/111 (51%)  
Query: 72 ESSQVTSQEEFWEFEFGREHRRQLWLEEEEMWQORQKXWALLEQEHQEKLRQWNLE 130  
E R++ E Q EE E RE R+L EEE++ Q+R+++ L QE KLR+ +  
Sbjct: 931 ERERKLREEEQQLRREEQLRREERARKL-REEEQQLQEREER-LRQERARKLREEE-Q 987  
Query: 131 DLAREQQRWVLEKEQESPRREPEQLGEDVERRITPTSRWDLKAEKSL 182  
L RE+Q +L +E++ RE EQL ++ E R R + E L  
Sbjct: 988 LLRREEQ---ELRQERDRKFREEEQQLQEREERLRQERDRKFREEERQL 1035  
Score = 101 (15.2 bits), Expect = 1.3e+00, P = 7.2e-01  
Identities = 33/108 (30%), Positives = 56/108 (51%)  
Query: 72 ESSQVTSQEEFWEFEFGREHRRQLWLEEEEMWQORQKXWALLEQEHQEKLRQWNLE 131  
E R++ E Q EE+ R+ R + EEE++ +Q +++ L QE KLR+ E  
Sbjct: 841 ERERKLREEEQQLRQEEQLRQERARKLREEEQQLRQEEQ---LRQERDRKLREE--EQ 895  
Query: 132 LAREQQRWVLEKEQESPRREPEQLGEDVERRITPTSRWDLKAE 179  
L R+++ +L +E++ RE EQL ++ E R R L + E  
Sbjct: 896 LLRQEEQ---ELRQERDRKLREEEQQLQEREERLRQERERKLREE 940  
Score = 99 (14.9 bits), Expect = 2.0e+00, P = 8.7e-01  
Identities = 32/97 (32%), Positives = 50/97 (51%)  
Query: 72 ESSQVTSQEEFWEFEFGREHRRQLWLEEEEMWQORQKXWALLEQEHQEKLRQWNLE 131  
E R+ E Q EE E R L EEE Q +++ L QE + KLR+ E  
Sbjct: 578 EKRRQERERQYRELEELRQEEQLRDRKLREEEQQLQEREERLRQERERKLREE--EQ 635  
Query: 132 LAREQ-----ORRWVLEKEQESPRREPEQLGEDVERRI 165  
L R++ Q R +L +E++ RRE +L ++ ER++  
Sbjct: 636 LLRQEEQLRQERERKLREEEQQLRREEQLRQERERKL 674  
Score = 99 (14.9 bits), Expect = 2.0e+00, P = 8.7e-01  
Identities = 34/111 (30%), Positives = 58/111 (52%)  
Query: 67 KQLSLESSQVTSQ--EFPWEFEFGREHRRQLWLEEEEMWQORQKXWALLEQEHQEKL 124  
++L E R++ E Q +E EE R+ R + EEE++ +Q +++ L QE + KL  
Sbjct: 664 QELRQERERKLREEEQQLQEREERLRQERARKLREEEQQLRQEEQ---LRQERERKL 720  
Query: 125 RQWNLEDLAREQQRWVLEKEQESPRREPEQLGEDVERRITPTSRWDLK 177  
R+ + L RE+Q L +E++ RE EQL ++ E R + L +  
Sbjct: 721 REEE-QLLRREEQL---LRQERDRKLREEEQQLQEREERLRQEREQQLR 768  
Score = 98 (14.7 bits), Expect = 2.6e+00, P = 9.2e-01  
Identities = 37/146 (25%), Positives = 77/146 (52%)  
Query: 20 EPVLLPVDRFPKMRFPVAESLGKXKXQEDYFQGGGLQKFMCSKQLSLESSQVTS 79  
E LL ++ ++ ER + E +E+ ++ K +QL + +++  
Sbjct: 655 EQLLRREEQLRQERERKLREEEQQLQEREERLRQERARKLREEEQQLRQEEQLRQ 714  
Query: 80 ESQEEFWEFEFGREHRRQLWLEEEEMWQORQKXWALLEQEHQEKLRQWNLE-LAREQOR 138  
E + + EEE + +RR+ L +E ++ +++ LL++ +E+LR+ E L RE+ R  
Sbjct: 715 ERERKLREEE-QLLRREEQLRQERDRKLREEEQQLQEREERLRQEREQQLRERDR 772  
Query: 139 RWVLEKEQESPRREPEQLG-EDVERRI 165  
++ E+Q RE E+L ++ ER++  
Sbjct: 773 KP--REEEQQLQEREERLRQERERKL 798  
Score = 97 (14.6 bits), Expect = 3.3e+00, P = 9.6e-01  
Identities = 38/129 (29%), Positives = 63/129 (48%)  
Query: 72 ESSQVTSQ--EFPWEFEFGREHRRQLWLEEEEMWQORQKXWALLEQEHQEKLRQWNLE 129  
E R++ E Q +E EE R+ R + EEE++ +Q +++ L QE KLR+  
Sbjct: 817 ERERKLREEEQQLQEREERLRQERERKLREEEQQLRQEEQ---LRQERARKLREE-- 871  
Query: 130 DLAREQQRWVLEKEQESPRREPEQLGEDVERRITPTSRWDLKAEKSLVPAPSRT 189  
E L R+++ +L +E++ RE EQL E+ + R R L + E L+



Sbjct: 872 EQLLRQEEQ---ELRQERDRKLREEEQLLRQEEQEL--RQERDRKLREE-QLLQSEEE 925

Query: 190 QSAHQSRPHL 200  
+ O R L

Sbjct: 926 RLRRQERERKL 936

Score = 96 (14.4 bits), Expect = 4.1e+00, P = 9.8e-01  
Identities = 41/132 (31%), Positives = 69/132 (52%)

Query: 46 KKKDOEDYFQGGGLQI-KFHCSKQLSLESSQVTSSEQEPWEEFGRMRRLWLEEE 104  
+++ QE F + Q+ + ++OL E S Q E + E+ G+ R QL +EE

Sbjct: 473 RERRQELQFLEEEQLRQERARQQLQEEQSFGEDRRRRRQQRPGQYWRWL---QEE 529

Query: 105 MWQRQKRWALLEQEHQELRQWNLLEDLAREQQRWVLEKEQSPREPEQLGEDVERR 164  
++R +A Q QE+LR+ E+L RE++R+ E+E+E E O ED +RR

Sbjct: 530 AQRRHTLYAKPGQ--QEQLREE--ELQREKRRQ----EREREYREEELRQEEDEKRR 581

Query: 165 IFTPTSRWRDLEK 177  
++R+LE+

Sbjct: 582 RQERERQYRELEE 594

Score = 96 (14.4 bits), Expect = 4.1e+00, P = 9.8e-01  
Identities = 35/138 (25%), Positives = 76/138 (55%)

Query: 28 DRFPFKWEPVAESL-GHKDKQEDYFQGGGLQIKFHCSKQLSLESSQVTSSEQEPW 86  
+R++ + E E L R +++E Q+ + ++ L Q+ + ++E

Sbjct: 586 ERQYRELEELRQEEQLDRKLREEEQQLQREERLRQERERKLREEEQLLRQEEQE-L 644

Query: 87 EEEFGRMRRLWL---EEEMWQRQKRWALLEQEHQELRQWNLLEDLAREQQRWV 143  
+E R++R + L EE+E+ Q+R++K L +E Q L++ E L R+++ R +L

Sbjct: 645 RQERERKLREEEQLLRQEEQLRQERERK---LREEEO-LQREERERLRQERAR--KL 698

Query: 144 EKQESPREPEQLGEDVERRI 165  
+E++ R+E ++L ++ ER++

Sbjct: 699 REEEQLRQEEQLRQERERKL 720

Score = 95 (14.3 bits), Expect = 5.2e+00, P = 9.9e-01  
Identities = 59/282 (20%), Positives = 121/282 (42%)

Query: 20 EPVLLPVDRRFPKWRPVAESLGHKDKQEDYFQGGGLQIKFHCSKQLSLESSQVTS 79  
E LL ++ ++ ER + E + +E+ ++ K +OL + +++

Sbjct: 655 EEQLLRREEQLRQERERKLREEEQQLQREERLRQERARKLREEEQLLRQEEQELRQ 714

Query: 80 ESQEPWEEFGRMRRLWLEEEEMWQRQKRWALLEQEHQELRQWNLLED-LAREQQR 138  
E + + EEE + +R+ L +E ++ ++ L++ +E+LR+ E L RE+ R

Sbjct: 715 ERERKLREEE--QLLRREEQLRQERDRKLREEEQLLQSEEEERLRQEREQLRAREDR 772

Query: 139 RWVLEKEQSPREPEQLG-EQVERRIITPTSRWRDLEKAEKLSLVPAPSRQSAHQ--S 195  
++ E+EQ RE E+L ++ ER++ ++ E+ L + + Q

Sbjct: 773 KF--REEEQQLQREERLRQERERKLREEEQQLQREERLRQERERKLREEEQLLQ 830

Query: 196 RRPPLMSPSTQOPALGQRPMSSVEFTYRPRTRVPTPKKSASFVVTGTSIRALTWPS 255  
R R + ++ L++ + E R R ++ +R+

Sbjct: 831 EREERLRQERERKLREEEQLLRQE-EQLRQERARKLREEEQLLRQEEQLRQERDRK 889

Query: 256 LQISFANIKKVVYHNDIEAQRK---NLQLSSEELALPHYLRKAL 299  
L+ +++++ + E RK QLL E E RL R + L

Sbjct: 890 LREEEQLLRQEEQLRQERDRKLREEEQLLQSEEEERLRQERERKL 936

Score = 94 (14.1 bits), Expect = 1.1e+00, P = 6.8e-01  
Identities = 35/116 (30%), Positives = 59/116 (50%)

Query: 72 ESSQVTSSEQEPWEEFGRMRRLWLEEEEMWQRQKRWALLEQEHQEK-----L 124  
E +R++ E Q EE+ R+ R + + EEE++ Q+R+++ L QE R L

Sbjct: 977 ERARKLREEEQLLRQEEQLRQERDRKFRREEQQLQREEE--RLRQERDRKFRREEQQL 1035

Query: 125 RQWNLLEDLAREQQRWVLEKEQSPREPEQLGEDVERRITPTSRWRDLEKAEKLSL 182  
R+ LE+ R+++ R +LE EQ +E +OL R F + R ++ E L

Sbjct: 1036 RQEELEEFQERDRKFRLE-EQIRQEEERQLRQERDRKFRREEQQRARRQERQQL 1092

Score = 94 (14.1 bits), Expect = 1.1e+00, P = 6.8e-01  
Identities = 51/166 (30%), Positives = 76/166 (45%)

Query: 67 KQLSLESSQVTSSESQ--EEPWEEFGRMR-RQWLEEEEMWQRQKRWALLEQEHQEK 123  
++L E R+ E Q +E EE R+ R R+L EEE++ + Q+ L QE+

Sbjct: 1250 QELRRERDRKFRREEQQLQREERLRQERARKLREEEQLLFEEQEEQRL----RQER 1305

Query: 124 RQWNLLED-LAREQQRWVLEKEQSPREPEQLGEDVERRITPTSRWRDLEKAEKLSL 182  
R++ E+ ARE++ R +LE+E R+E EQ R F R E+ E

Sbjct: 1306 DRYRAEEQFAREESR--RLREL---RQEEQRARRERARKFRREEQLRQEE-EQRR 1359

Query: 183 VPAPSRTOAHOSRRPHLPMSPSTOOPALGKORPMSSVEFTYRPRTRVP 232  
R QSRRL P T+Q A R E+ R++ P  
Sbjct: 1360 AOLREROFEDQSRAQVL--EPGTRQFARVPVRSSPLYEQORSQYRP 1407  
Score = 93 (14.0 bits), Expect = 8.3e+00, P = 1.0e+00  
Identities = 41/145 (28%), Positives = 72/145 (49%)  
Query: 28 DRFFPKWERPVASLGHKQKQDQDYFQGGLOTKFNCSKQLSLESSROVTSQEEPM- 86  
+RR ++ E+ E + + Q + + Q + L R + QE+ +  
Sbjct: 408 ERRQQRERERELEQARRQQQWQAEESERRRQ-RLSARPSLRERQLRAEERQEQQRFR 466  
Query: 87 -EEFGREMRQL-WLEEDMWOORQKQKQWLEQ--HQEKLQWNLLEDLAREQQRWVQ 142  
EEE RE R++L +LEEE Q+R++ L E++ +++ R+ ++ Q RW Q  
Sbjct: 467 EEEQQRERERQQLQLEEEQQLRRRAQQLQEDSFOEDREARRAQEQRPQOTWRW-Q 525  
Query: 143 LEKEQESPRR----EP---EQLGEDVE 162  
L++E + R +P EQL E+ E  
Sbjct: 526 LQEEAQRNHTLYAKPGQEQLEEE 552  
Score = 91 (13.7 bits), Expect = 2.4e+00, P = 9.1e-01  
Identities = 38/110 (34%), Positives = 57/110 (51%)  
Query: 72 ESSROVTSQEEPMEE-EPGREMRQLWLEEDMWOORQKQKQWLEQHQEKLQWNL- 129  
E R++ E Q EE E RE R+L EEE+ Q+R+++ L QE KLR+  
Sbjct: 931 ERERKIREEQQLRREQLRRERARKL-REESQLOERFEE-RLRQERARKLREEQQL 988  
Query: 130 -----EDLAREQQRWVQLEKEQESPRPEQLEDVERRIFTPTSRWRDLEKAE 180  
++L +E+ R++ E+EQ RE E+L R F R L+ EL  
Sbjct: 989 LRREQQLRQERDRKF--REEQQLQREERLQERDRKFRFEEER--QLRQEL 1040  
Score = 89 (13.4 bits), Expect = 2.2e+00, P = 8.9e-01  
Identities = 35/138 (25%), Positives = 65/138 (47%)  
Query: 82 QEEPWEEFGREMRQLWLEEDM--WQORQKQWLEQHQEKLQWNLLEDLAREQQR 139  
Q E++ E+R + + +E E WQ++++ L E+ E Q K R+ + +R+ + +  
Sbjct: 111 QNRQEDQRRFELRDQFQDEPERRRWQKQEQERELAEKEQKRRERFEQYRSQYRDK 170  
Query: 140 WVQLEKEQ-ESPRPEQQL----GDEVERRIFTPTSRWRDLEKAE LVPAPSRTOAHQ 194  
+L++++ E R E EQL G D E F + R E+ EL Q +  
Sbjct: 171 EQRLQRELEERAEELARRKGRDAEE--FIEEQLRREQQLR-ELREEQORRE 227  
Query: 195 SRRPHLPMSPSTOOPALGQR 215  
R H ++ L +R  
Sbjct: 228 RREQERALEEEELRQR 248  
Score = 50 (7.5 bits), Expect = 2.2e+00, P = 8.9e-01  
Identities = 34/160 (21%), Positives = 67/160 (41%)  
Query: 325 RLQSLQEAINHVMQINKETEASYKAQLYIFLENDRL-QSLRLQAMTDKQGLEEKHRE 383  
R + R+E Q+ +E E + + LE +R Q LR + + + + E+ R  
Sbjct: 245 QRARWKEPREQQLARELEKIRERQR--LEQERREQQLAREQLEQERREQQLAR 301  
Query: 384 CLSSMVTMFKLQLENNVHLNIP-EVTSFKPKCKLPASPRHIPSQPTYKOPFLSRH 442  
L + +L E + E + K+L R R + + L+  
Sbjct: 302 ELEIRERERQLEQERREQRLEQERREQQLKRELEIRERERQLEQERREQQLAEV 361  
Query: 443 ACVPLQARQOQKQHEAVNKEVASSYAEKXTPASLPRDQ 484  
+ AR++G+ + W+ ++ S + A + K + S PR Q  
Sbjct: 362 R----EQAREGESLTRWQROQLESGARQSKV-YSRPRQ 398  
Score = 40 (6.0 bits), Expect = 1.9e-01, P = 1.7e-01  
Identities = 32/115 (27%), Positives = 47/115 (40%)  
Query: 276 RKNLQLLSESESLRPHYLRKAL--ELTTTTELGAIRLQYLCHRYIFYRRL-OSLRQE 332  
R+ QLL E E RL R+L E E E LR Q R+ +L Q +E  
Sbjct: 959 REEQQLQREERLRRQERARKLREEQQLRREQQLR-QERDRFREDQLQEREKE 1017  
Query: 333 AINHVIQI--MKETEASYKAQLYIFLENDRLQSLRLQAMTDKQ-KGLEEKHRE 383  
+ + + +E E + Q L F + DR L Q +R+ K L + R+  
Sbjct: 1018 RLARQERDRKFREERQLRQLEEQFRQERDRFLEEQIRQEKERQLRQERD 1073  
Score = 37 (5.6 bits), Expect = 1.6e+00, P = 7.9e-01  
Identities = 27/108 (25%), Positives = 43/108 (39%)  
Query: 276 RKNLQLLSESESLRPHYLRKAL--ELTTTTELGAIRLQYLCHRYIFYRRLQSLRQE 332  
R+ QLL E E RL R+L E E LR Q K R + L QE  
Sbjct: 775 REEQQLQREERLRRQERARKLREEQQLQREERLRRQERARKL--REEQQLQE 831  
Query: 333 AINHVIQINKETEASYKAQLYIFLENDRLQSLRLQAMTDKQ-KGLEEKHRE 383  
+E E + + + E L+ R+ + + + L + + +E  
Sbjct: 832 REERLRRQERARKLREEQQLRQE-QELRQERARKLREEQQLRQE 881

## Pedant information for DKFZphtes3\_50n23, frame 1

## Report for DKFZphtes3\_50n23.1

```
[LENGTH]      499
[MW]           50885.69
[pI]           9.67
[KW]           A1.1_Alpha
[KW]           LOW_COMPLEXITY    10.42 %

SEQ  MTVRSRVADVFGSKOTESLEPVLLPLVDRRFKKWERPVAESLGHKDKQEDYFQKGGLO
SEG  .....
PRD  ccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc

SEQ  IKFHCCKQLSLESSRQVTSSESQEEPMWEEFGEMRRQLWLEECMMQQRQKKWALLEQEH
SEG  .....
PRD  #####cccccccccccccccccccccccccccccccccccccccccccccccccccccccccc

SEQ  QEKLRQWNLEDLAREQQRMMVOLEKEDQSPRPEQLGEDVERRIPTPTSRWRDLEKAEI
SEG  .....
PRD  hhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh

SEQ  SLVPAPSRQSAHOSRRPHLPMSPTQOPALGQRPMSSEFTYRPRTRRVPTKPKKSAS
SEG  .....
PRD  hcccccccccccccccccccccccccccccccccccccccccccccccccccccccccc

SEQ  FVVTGTSIRRLTWPQLQSPANIKKKVYHMDIAQRKMLQLLSESELRLPHYLRSKALE
SEG  .....
PRD  #####cccccccccccccccccccccccccccccccccccccccccccccccccccccccccc

SEQ  LTTTINELGALRLQYLCHKYIFYRRLQSLRQEAINHVOIMKETEASYKAQNLVIFLENID
SEG  .....
PRD  hhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh

SEQ  RLQSLRLQAWTDKQKGLKRECLSSMTMFPKLQLEWVHLNIPVTSPPKPKCKLPA
SEG  .....
PRD  hhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh

SEQ  ASPRHIRPSGPTYKOPFLSRHRACVPLQMARQCKQMEAVWKTEVASSSYAIEKKT PASL
SEG  .....
PRD  ccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc

SEQ  PRDQLRGHPDIPRLILDV
SEG  .....
PRD  ccccccccccccccccccc
```

(No Prosite data available for DKFZphtes3\_50n23.1)

(No Pfam data available for DKFZphtes3\_50n23.1)

DKFZphtes3 6b21

group: testes derived

DKFZphtes3 6b21 encodes a novel 781 amino acid protein without similarity to human KIAA0256 gene product.

No informative BLAST results: No predictive prosite, pfam or SCOP motif.

The new protein can find application in studying the expression profile of testis-specific genes.

similarity to KIAA0256

complete cDNA, complete cds, EST hits

Sequenced by BMFI

Locus: /map="J56.3 cR from top of Chr9 linkage group"

Insert length: 3360 bp

Poly A stretch at pos. 3314, polyadenylation signal at pos. 3300

```
1  GGCAAGCCGA  CGGCCGCTG  CTGGCCTCG  TGACGGGGC  TCCTCCGCG
51  CTCGCGCAT  GCGTCGGAG  GGGCCGCGG  AGCCGAAG  CGAGGGCATC
101  AAGTATACG  CAGATGTCG  ACCATTTGC  CCGCATTTG  CGGGCTGCA
151  TGTGGCATG  TTAGAGTCG  CAGAAAGATG  TGTCTTCCC  AGCTCTGAG
201  CCACATACT  TCGTTTGTG  CAGGAACCC  CAGTGACAG  AATGTTTACT
251  CAGTGGCTG  CTGCCAGTAT  CTTTATACC  AACCCAGTG  TTACGAGGT
301  TTCAACAGC  TGAAGCATG  AATGAGAGC  ACATGGCTC  TCCACAGA
351  AATGAAGCT  CTGTTTAA  AGAAACCTA  TGATGAGAA  AAAAGTATG
401  ATCAGCAAA  GTTTGACAG  GAAAGGGCT  ATGGAATAT  ATCATCTGAG
451  ATAAATACG  CTAGAGGTC  ACATCATTT  TGCATTAGG  CTGAGATAG
501  TTGAATACA  GATGTTTAC  ATAGCGAAG  AGACAGGAA  TCCAATGA
551  TTGCAAAAA  TGTATCTAC  TCCAACCTG  AGTTGAATT  TACCACACT
601  GACTTTCCT  AACTGCAAG  TGCAGAGAC  AATATGTCG  AGATACAGAA
651  GCAACCCAG  TGGGAGCTG  TCCACTCTG  CTCTACGAC  ATTCTCTTC
701  TAGGAGAGT  AGTAAACCA  GCTGAGTGT  TATCAAGGG  TGAATATGT
751  GTGAAAAAT  ACCCAATGA  ATCTGTAAT  GCTAATGCC  CTACCAATT
801  TCCTTCAAT  ACAAGAGAG  TATCTTGGC  ACCAATGGT  TATGTTGTC
851  GACAGACAT  ATCTACAGA  CTGTCAGAG  CCGCTAAAA  TGTACTTCT
901  ATGATAAAT  TAAAGACAT  TCGTTCATC  GCAGATCTA  AAAATGTTG
951  TATACCATC  TCTGAGCTT  TATCTTGGG  TCCTTCTAC  AACAAGAAA
1001  AACACATTAT  TCATCTTAC  CAAAAGTCT  AAGCATACA  AGTAGTGAC
1051  CTTGAGCAAA  ATGAAGCTC  AAGAAAGAT  AAGAAAAAG  AAGAAAAATC
1101  TACATCAAA  TATGAATCC  TGACAGTCA  AGAGCTCCA  AGATTTGAG
1151  ATGGCGGAG  ATTTCCCAAC  CTGGCAGTG  CATCTGAAG  AAGAGACAG
1201  ATAGAGACAC  CGAAATTTC  ATCTAAGCA  CAGCCACAG  ATAATTTAA
1251  AAATAATGT  AAGAAGAGC  AGCTTCCAG  GCAGTTGAG  TTGGGGGCA
1301  TCCTGACAC  CTTGGAGAG  AGCAGACT  CTCAGATCC  AAGCAGTCC
1351  TCCAACCCG  TGGTAGTCT  AGTTGAGCA  GTGCCATCC  TTCCAAAGA
1401  ATGTGCATC  GGGGAGAG  GCGCGCGAT  GAGTCAATG  AAGCCCGCG
1451  ACAATCCCT  GGACTCCAG  GCGCCACTG  TGAAGAAAG  GAAGCAGAG
1501  GAGATCCCA  AGCCCAAGA  GCGCAACTC  CTGAAGAAG  TTATTTGAA
1551  AGAATCCCA  GAGAGAAAG  AGCCTTCCA  AGAAATGCT  GTAGTCCAG
1601  CTTTACCAG  TGATGACAC  CAAGATGGG  AGAGTGTGG  TGATGACAG
1651  TTTCCCGAG  AGGCAGAGC  GTCAGGGCA  GAGGGATGG  ACGACTGAT
1701  CTCCACTCT  TCGGTTGAG  ACAAGTCTG  AGAGCCACA  GGCACAGC
1751  TCCAGAGGA  CACAGAGCC  TCCCACTTG  CTCCCAATC  CACCACCTT
1801  CTAAGATCC  ACAGCGCAG  ATTCAGGAT  TACTGAGCC  AGATGCTT
1851  TAAAGAGTG  GATGCTTGT  TTACCGACT  ACTCAAGAA  CTGTCGCTT
1901  TCCAAGACG  TATGTACAG  AAAGATCCG  TCAAGCCCA  GACTAAAGT
1951  CGACTTGTG  TGGGTTGAG  GAGGCTTCT  TCAAGCTCA  AGCTCAAAA
2001  ACTGAATGT  GTCATTATT  CTCCCACTG  TGAGAAGAT  CAGTCAAA
2051  GTGGGCTGA  TGACACTTG  CACACAATA  TTGATTATG  CTGTGAGCA
2101  AACATTCCCT  TTGTGTTTG  TCTAACCGG  AAAGCTTGG  GAGCAGTTT
2151  GAATAGGCA  GTTCTCTCA  GTGCTGGGG  GATCTTCCG  TATGATGGG
2201  CCCAGATCA  GTTCCAGAG  ATGTTGAGC  TGACAGTGG  GCGCCGACG
2251  CGGTACAGA  CCAATGCTG  GAATGTGAG  CAGGAGCTG  TGGGAGAGC
2301  CAGGCGTCAG  GCACCTCCA  CCTAGCCAG  ACAGGGCCC  AGCTGCCCT
2351  CAGAGATGG  CCCCCAGC  CTGAAGAAA  AAGAGAGCC  ACATCAAT
2401  GAAATCTGA  AAAACATCT  GGAAGCATC  AGTGGATGA  CCTGGAGCT
2451  AGAAGATCC  TTGAGGCTT  CAACCTCTA  AATGATGAT  TTGAATTTAT
2501  GAGAGTTCT  GCTGTGTGT  CTGTATTTG  GGTAAAGAG  GGAGTCTGA
2551  AAAAGACTT  GGGGCTTTT  CTCTGTTTG  TATGACAT  GTAAATTTG
2601  TAACCTTGA  ATCTGGAAT  TGATCAGAT  TAAAGGGCA  ATGACAGCT
2651  GTCTGAGGC  GTTCAGTGT  GCGAGGCTG  TTAAGGTCA  CTCAGATGT
```

```
2701 CAGGCTGTAA TCTTCTCTAA AAGCCTGGTT ATACAGCTCT GGCTTTCTGA
2751 GCACACTAGG GATCTGGAAA ATACTGGAAA ATGTGATCT TAGAATACTT
2801 TGGCTGCTAA GGAAGCTTCC TCTCCATTGC AGAATAGCTG AGCCAAGTGA
2851 GTGAGTTTTC AGAAGCAGG TGGTGAGCTC CTGCTGCTG GAGGTTUCCA
2901 TGGAGGGKCA TTCTGCCCC GCAAGAGCAC CTCTGTGAG GAGGCTACTT
2951 GGCAGAAGGG TGCAGGGCTG CTGGTGTGAG AGCAAGAGGG CTACAGGGAA
3001 AGGGCCCTTT CTCAGGGGAT GTAGCTTTTT TAAAGATT TGGGACACTT
3051 GGAGGATTTC CTAAATGAG CCTCAGAGG AAAATTGUTT TTCTAACCTG
3101 TGACTTTTTT AAATGAATTA TTCTTTTTCAG TCTTTATTTT TCAAGAAC
3151 AATGTGTATT GAAGTACCTA GATTGTGTTG ATAATCAACA AATCTTCTCT
3201 TTTTCAATGA ACATATTCTG AATGTGTTT CTGCTTTAGA CCAGGAGGAC
3251 AGAGTTTGCT TTCATATTTT CCTGTAAAGT AAGAGGGCTT ATTATTTTTA
3301 AATAAAGAGT AATTATTAAA AAAAAA AAAA AAAA
3351 AAAAAA
```

## BLAST Results

Entry HS773347 from database EMBL:  
human STS W1-18160.  
Score = 813, E = 2.9e-30, identities = 167/171

## Medline entries

No Medline entry

## Peptide information for frame 1

ORF from 157 bp to 2499 bp; peptide length: 781  
Category: similarity to known protein

```
1 MVRVLRSMCL POLCSHILSV CSCTISDRNV YSPGSOYLY NQPSYRGFO
51 TVKRRMENTC PLPOEMHALF KRRTYDEKAT VQQRTOSEK ADOYISSEIK
101 SARGSHHLSI YAENSLKSDG YHKRTDRKSR IIAKNVSTSK PEFEFTTLDF
151 PELQGAENNM SEIQKQPKWG PVHVSVDIS LLREVVKPAA VLSKEIVVK
201 NNPNESVTAN AATNSPSCYR ELSWTFNGYV VQRTLSLELS AAPNVTSMI
251 NLKTIASSAD PRNVSTPSE ALSOSPVMK EKHIIHPQK SKASQGSOLE
301 QNEASRNKK KKEKSTSKYE VLTQEPPI EDAEEFNL VASERRDRIE
351 TPRFOSKQOP QDNFNHVKK SOLPVOLDLG GMLTALEKQ NSQAKQSSK
401 PVVSVGAVP VLSKECASOE RGRMSQKRT PINPLOSSAP LMRKQOREI
451 PKARKFTSLK KILKEKDER KQLQENAVS PAFTSDDYD GEGGGDOOP
501 EQNELSGPEG NDELSTPSV EDKSEEPPT ELORDTEASH LAPNHTTFPK
551 IHSRRFRDYC SQMLSKVEDA CVTDLKELY RFQDRHYQKD PVKATKRRAL
601 VLGLREVLRH LKLKELKCVI ISPNCEKIQS EGGLODTLHT IIDVACEQNI
651 PFVFLNRKA LORSLEKAVP VSPVGIPTD GAGDOYHNV ELTVAAQRY
701 KTNLENVOOE LVGEPRQAP PSLPTGQPSK PAEDGPPALK EKEEPHYIEI
751 WKKHLEAYSG CTLELESLE ASTSQMNIN L
```

## BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DRFZptes3\_6b21, frame 1

SWISSPROT:Y256 HUMAN HYPOTHETICAL PROTEIN KIAA0256., N = 1, Score = 786, E = 3.6e-78

TREMBL:PFNAL3P3 15 gene: "MAL3P3.15"; Plasmodium falciparum MAL3P3, N = 2, Score = 161, E = 5.1e-10

TREMBL:RNNFLH.1 Rat heavy neurofilament subunit (NF-H) mRNA, 3' end., N = 1, Score = 150, E = 9.1e-07

>SWISSPROT:Y256 HUMAN HYPOTHETICAL PROTEIN KIAA0256.  
Length = 635

## HSPa:

Score = 786 (117.9 bits), Expect = 3.6e-78, E = 3.6e-78  
Identities = 190/424 (44%), Positives = 263/424 (62%)

Query: 369 KESQLPVOLDLGGMLTALEKQHSQAKQ---SSKPVVSVGAVPVLSKECASGERGRMS 426  
EK++ PVOLDLG ML ALEK+Q + A+Q +++P+ +V + + + S  
Sbjct: 16 KKNKTPVOLDLGGMLAALKKQQAQKARQITNTRPLSYTVVTAASFHTKOSTNRKPLTKS 75

Query: 427 Q-MKTPHNPLDSSAPLMKKQKQREIPKAKKPTSLKKIILKERQERKQRLQENAVSPAFTS 485  
Q T N +D ++ KKGK++EI K K+PT+LKK+ILKER+E+K RL + S  
Sbjct: 76 QPCLTSFNSVDIASSKARKGKEIKAKLRPTALKKVLKEREKGRITVD--HNLGGS 133

Query: 486 DOTQDGESGGDDOTPEQALSGPEQMDLISTPSVEDKSEEPG--TELQDTEASML-- 541  
++ + D P++ G+ + S S+ S+ P T + + + AS  
Sbjct: 134 EEPTMHLOFIDDLPOEIVSQEDTGLS-NPSDTLSPASQNSPYCHTVPVSGSPASSGIG 192

Query: 542 APN-HTTFPKIHSRRFRDYCSQMLSKVDACVTDLLKELVRFODRMVYQKDPVAKTKARL 600  
+P +T KHS+RFR+YC+Q+L KE+D CVT LL+ELV FQ+R+YQKDPV+AK +RRL  
Sbjct: 193 SPMASSSTITKINSKRFREYCNQVLCKEIDECVTLLQELVSFQERIYQKDPVRAKARRL 252

Query: 601 VLGLREVVLKHLKLLKCVIISPNCETIOSKGLD+L+ +I A EQ IPVFAL RKA 660  
V+GLREV KH+KL K+KCVIISPNCETIOSKGLD+ L+ +I A EQ IPVFAL RKA  
Sbjct: 253 VNLGREVTKHKLKIKCVIISPNCETIOSKGLDEALYNVIAAREQEIIPVFALGRKA 312

Query: 661 LGRSLNKAVPVSVVGIFSYDGAQDQTHMVELTVAARQAYKTMLEHVQOELVGEPR-- 717  
LGR +NK VPVSVVGIF+Y GA+ F+K+VELT AR+AYK M+ ++OE E  
Sbjct: 313 LGRGVNKLVPVSVVGIFNYFGAELFNKLVELTEARAYKDMVAAMEQEAELKNVK 372

Query: 718 QAPPSLP-TQGPS-----CPAEDGPPALKEKEEPHYIWKHLEAYSGCTL---ELE 766  
+ P + ++ PS C P + E E Y W+ +E G E E  
Sbjct: 373 KVPNHMGHSRNPAAASAIAPCVISEP--LSEVNEKEYETNWRNMVETSDGLEANEKE 430

Query: 767 ESLEASTSQ 775  
S + STS+  
Sbjct: 431 VSCKHSTSE 439

Pedant information for DKF2phtes3\_6b21, frame 1

Report for DKF2phtes3\_6b21.1

[LENGTH]	781
[HW]	87393.44
[PI]	8.94
[HOMOL]	SWISSPROT:Y256_HUMAN HYPOTHETICAL PROTEIN KIAA0256. 4e-75
[PROSITE]	MYRISTYL 4
[PROSITE]	AMIDATION 1
[PROSITE]	CAMP_PHOSPHO_SITE 3
[PROSITE]	CK2_PHOSPHO_SITE 16
[PROSITE]	TYR_PHOSPHO_SITE 4
[PROSITE]	PKC_PHOSPHO_SITE 16
[PROSITE]	ASN_GLYCOSYLATION 6
[KW]	Alpha_Beta
[KW]	LOW_COMPLEXITY 8.45 %

SEQ MVRVLRSMCLPQLCSHILSVCSGTTSDRWVSVVPGSYLYNQPSCYRGFQTVKRRNEMTC  
SEG .....  
PRD ccc

SEQ PLPQEMKALFKKTYDEKTYDQKFDSEADGTISSEIKSARGSHHLSIYAENSLKSDG  
SEG .....  
PRD ccc

SEQ YKRTDRKSRIIAKNVSTSKPEFTTLOFPELQGAENHSEIQKQPKWGPVHSVSTDIS  
SEG .....  
PRD ccc

SEQ LLREVVKPAAVLSKGEIVVKNPNHESVTANAATNSPCTRELSTWPMGYVVRQLSTELS  
SEG .....  
PRD hhhhhhhcc

SEQ AAFKNVTSIMINKTIASSADPNVSI PSSEALSSDPVYNKEKHIIHPTQKSKASQGSLE  
SEG .....  
PRD ccc

SEQ QNEASRRNKKKKEKSTSEVLTVEPPRIEAEFPNLAVASERRDRIDTQFKQSKQQP  
SEG .....  
PRD hhhcc

SEQ QDNFKNNVKSQLPVOLDLGGMLTALEKQHSQAKQSSKPVVSVGAVPVLSKECASGE  
SEG .....  
PRD ccc

```
SEQ RGRNSQKTPHNFLDSSAPLMKKGKQREI PKAKKPTSLKKI ILKERQERKQRLQENAVS
SEG .....
PRD chhhhhhccccccccccccchhhhhhhhhhhccccchhhhhhhhhhhhhhhhhhhhhcc

SEQ PAFTSDDTQDGESGGDDQFPEQALSGFEGHDELSTPSVEDKSEEPGTQLQRTAEASH
SEG .....
PRD cccccccccccccccccchhhhhhhcccccccccccccccccccccccccccccccc

SEQ LAPNHTTFPKIHSRRFRDYCSQHLSEVDACVTLKELVRFQDRNYQKDPVKATKRRL
SEG .....
PRD cccccccccccccchhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh

SEQ VIGLAEVLKHLKLLKCVIISPNCEKIQSKGGLDOLTHTIIDYACEQNIFFVFNLRKA
SEG .....
PRD hhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh

SEQ LGRSLKAVPVSVVGIFSYDGAQQDFHMYELTVAARQAYKTMLENVQQLVGEPRPOAP
SEG .....
PRD cccccccccccccccccchhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh

SEQ PSLPTQGPSCPAEDGPPALKEKEEPHYIETWKKHLEAYSGCTLEESLEASTSQENHLN
SEG .....
PRD cccccccccccccchhhhhhhccccccccccccchhhhhhhhhhhhhhhhhhhhhhhhhhh

SEQ L
SEG .
PRD c
```

## Prosite for DKFZpites3\_6b21.1

PS00001	135->139	ASN_GLYCOSYLATION	PDOC00001
PS00001	159->163	ASN_GLYCOSYLATION	PDOC00001
PS00001	204->208	ASN_GLYCOSYLATION	PDOC00001
PS00001	245->249	ASN_GLYCOSYLATION	PDOC00001
PS00001	263->267	ASN_GLYCOSYLATION	PDOC00001
PS00001	544->548	ASN_GLYCOSYLATION	PDOC00001
PS00004	71->75	CAMP_PHOSPHO_SITE	PDOC00004
PS00004	423->427	CAMP_PHOSPHO_SITE	PDOC00004
PS00004	454->458	CAMP_PHOSPHO_SITE	PDOC00004
PS00005	26->29	PKC_PHOSPHO_SITE	PDOC00005
PS00005	51->54	PKC_PHOSPHO_SITE	PDOC00005
PS00005	88->91	PKC_PHOSPHO_SITE	PDOC00005
PS00005	101->104	PKC_PHOSPHO_SITE	PDOC00005
PS00005	115->118	PKC_PHOSPHO_SITE	PDOC00005
PS00005	125->128	PKC_PHOSPHO_SITE	PDOC00005
PS00005	138->141	PKC_PHOSPHO_SITE	PDOC00005
PS00005	288->291	PKC_PHOSPHO_SITE	PDOC00005
PS00005	305->308	PKC_PHOSPHO_SITE	PDOC00005
PS00005	316->319	PKC_PHOSPHO_SITE	PDOC00005
PS00005	343->346	PKC_PHOSPHO_SITE	PDOC00005
PS00005	351->354	PKC_PHOSPHO_SITE	PDOC00005
PS00005	398->401	PKC_PHOSPHO_SITE	PDOC00005
PS00005	458->461	PKC_PHOSPHO_SITE	PDOC00005
PS00005	553->556	PKC_PHOSPHO_SITE	PDOC00005
PS00005	596->599	PKC_PHOSPHO_SITE	PDOC00005
PS00006	24->28	CK2_PHOSPHO_SITE	PDOC00006
PS00006	74->78	CK2_PHOSPHO_SITE	PDOC00006
PS00006	139->143	CK2_PHOSPHO_SITE	PDOC00006
PS00006	146->150	CK2_PHOSPHO_SITE	PDOC00006
PS00006	193->197	CK2_PHOSPHO_SITE	PDOC00006
PS00006	257->261	CK2_PHOSPHO_SITE	PDOC00006
PS00006	297->301	CK2_PHOSPHO_SITE	PDOC00006
PS00006	317->321	CK2_PHOSPHO_SITE	PDOC00006
PS00006	323->327	CK2_PHOSPHO_SITE	PDOC00006
PS00006	384->388	CK2_PHOSPHO_SITE	PDOC00006
PS00006	484->488	CK2_PHOSPHO_SITE	PDOC00006
PS00006	493->497	CK2_PHOSPHO_SITE	PDOC00006
PS00006	506->510	CK2_PHOSPHO_SITE	PDOC00006
PS00006	519->523	CK2_PHOSPHO_SITE	PDOC00006
PS00006	640->644	CK2_PHOSPHO_SITE	PDOC00006
PS00006	702->706	CK2_PHOSPHO_SITE	PDOC00006
PS00007	581->588	TYR_PHOSPHO_SITE	PDOC00007
PS00007	740->748	TYR_PHOSPHO_SITE	PDOC00007
PS00007	740->748	TYR_PHOSPHO_SITE	PDOC00007
PS00007	73->82	TYR_PHOSPHO_SITE	PDOC00007
PS00008	93->99	MYRISTYL	PDOC00008
PS00008	155->161	MYRISTYL	PDOC00008
PS00008	380->386	MYRISTYL	PDOC00008

WO 01/12659

PCT/IB00/01496

PS00008	633->639	MYRISTYL	PDOC00008
PS00009	421->425	AMIDATION	PDOC00009

(No Pfam data available for DKF2phtes3\_6b21.1)



WO 01/12659

PCT/IB00/01496

DKFZphtes3\_6c11

group: signal transduction

DKFZphtes3\_6c11 encodes a novel 1025 amino acid protein with similarity to A. ambisexualis antheridiol steroid receptor.

The novel protein is a putative steroid receptor. It shares similarity with yeast YNL132w and contains the ATP/GTP-binding site motif A (P-loop) and RGD site, similar to the A. ambisexualis antheridiol steroid receptor.

The new protein can find application in modulating/blocking the expression of genes controlled by this receptor.

strong similarity to YNL132w

strong similarity to S.pombe/YDR9\_SCHPO, S.cerevisiae/YNL132w, C.elegans/F53A12.0

Sequenced by BMF2

Locus: unknown

Insert length: 3966 bp

Poly A stretch at pos. 3890, polyadenylation signal at pos. 3873

```

1 GCTGTGCCCT CTCTTCGGA GTTGTCCGT GCTCCACGT GCTTCCCTT
51 CTCCACTGGC TGGATCCCC CGGCTCGGG GCGCAGTAT AATTTTCAC
101 CATGCATCGG AAAAGGTGG ATACCGAAT CCGATCTCT ATTGAGATG
151 GAGTACCTGA GCGGCAAGA TCTCTCTTTG TTGTAGTTGG GGATCGAGGA
201 AAAGATCAGG TGGTAATCT TCATCACATG TTATCCAAAG CAACGTGAA
251 GGCTCGGCGT CTACTGCTGT GGTGTTTAA GAAAGAGCTG GGGTTAGCA
301 CTCACCGGAA GAAAGATAT CGACAGCTG AGAAGAAAT AAGAAATGA
351 ACATGAAAG TAAAGCAGA CGACCCCTTT GAATCTTCA TAGCAGCAC
401 AAACATTGCG TACTGCTACT ACAACGAGC CCACAAATG CTGGGCAATA
451 CCTTGGGCAAT GTGTGTGCTG CAGGATTTTG AAGCCTTAAC TCCAAACTTG
501 CTGGCCAGGA CTGTAGAAG ACTGCAAGT GGTGGGCTAG TGGTATCTCT
551 CCTAGCGAGC ATGAATCTAC TGAAGCAAT GTACACAGTG ACTATGGATG
601 TGCATTCCAG GTACAGAAT GAGGCCATC AGGATGTGGT GGGAGATTT
651 AATGAAAGGT TTATCTCTG TCTGGCTCT TGTAGAAGT GTCTGCTCAT
701 TGATGACCAG CTCACATCC TGCCCATCT CTCACAGTT GCACCAATGG
751 AGGCTCTGCC TCCCGAGCT CCGGATGAGA GTCTTGCTCC TTCTGATCTG
801 GAGCTGAGGG AGTTGAAGA GAGCTTCAG GACCCGAGC CTGTGGGTGT
851 GTTGTGGAGC TGCTGAAGA CTCTAGACCA GGCAGAGCT GTCTTGAAT
901 TTATCGAGGG CATCTCTGA AGACCTCTGA CGATCTCTG TGCATCTACA
951 GCTGCTCGAG GACGGGAAA ATCTCGAGCC CTGGGATTGG CGATTGCTGG
1001 GCGGCTGGCA TTTGGTACT CCAATATCTT TGTACCTCC CCAAGCCCTG
1051 ATAACTTCCA TACTCTGTTT GAATTTGTAT TTAAGGATT TGAATGCTCTG
1101 CAATATCAGG AACTCTGGA TTATGAGATT ATCCAGTCT TAATCTCTGA
1151 ATTAAAGAA GAGTGAATCA GAGTGAATGT ATTTGAGAA CACAGGAGAA
1201 CTATTCACTA TATACATCT GCAGATGCTG TGAAGCTGGG CCAGGCTGAA
1251 CTAGTTGTGA TTGATGAAGC TGCCGCCATC CCGCTCCCTT TGGTGAGAG
1301 CTACTCTTGG CCCTACCTTG TTTTCATGGG ATCCACCATC AATGCTGATG
1351 AGGGCACTGG CCGTCACTG TCCCTCAAGC TAATTCAGCA GCTCCGTCAA
1401 CAGAGCGGCC AGAGCCAGGT CAGCAGCACT GCTGGAATA AGACCAGGAC
1451 GACAGCCAGA TTGGCATCAG CGCGGACACT GCATGAGGTT TCCCTCCAGG
1501 AGTCAATCCG ATACGCCCTT GGGGATGAG TGAAGAAGTG GCTGAATGAC
1551 TTGCTGTGCG TGGATTGCTT CAGATCACT CGATATGCT CAGGCTGCCC
1601 CTTCGCTGAA GCTTGTGAAC TGTACTGTGT TAATAGAGAT ACCCTCTTTT
1651 GCTACCACAA GGCCTCTGAA GTTTCTCTCC AACGGCTTAT GGCCCTCTAC
1701 GTGGCTTCTG ACTACAAGAA CTCTCCCAAT GATCTCCAGA TGCTCTCCGA
1751 TGACCTCTGT CACAATCTT TCTGCTTCT GCTCTCTGT GCGCCCAACC
1801 AGAATGCCCT TCCAGAAGTG CTGCTGTGTA TCCAGTGTGT CTTTGAAGGG
1851 GAGATTTCTC GCCAGTCCAT CTGGAACAGT CTGTCTGAG GCAAGAAGGC
1901 TTCAGGGGAC CTGATTCCAT GGACAGTGTG AGAACATTTT CAAGATCCAG
1951 ACTTTGCTGG TCTGTCTGCT GGAAGGCTGG TTGCTATTGG TGTTCAGCCA
2001 GATTATCAAG GGATGGGCTA TGGCAGCCGT GCTCTGAGC TGCTGCAGAT
2051 GTACTATGAA GGCAGGTTTC TTGTCTGGA GGAAGAGGTC CTTGAGACAC
2101 CACAGGAAAT TCACACGTA AGCAGGAGG CTGTGAGCTT GTTGAAGAG
2151 GTCACTACTC CCGGAGAGA CTGCGCTCT TTAATCTCTA AATTAAGTA
2201 GAGGCTCTGC GAACGCTTG ATTACTGGG TGTTCTCTAT GGCTTGACCC
2251 CCAGGCTCTC CAAGTTCTG AAACGAGCTG GATTGTCTCC TGTTTATCTG
2301 AGACAGACCC CGAATGACCT GACCGGAGAG CACTGTGCA TCAATGCTGA
2351 GACGCTCACT GATAGGATG AGCTGAGCA GGAAGGCTGG CTTCAGGCT
2401 TCTGGAAGAA TTTCGAGGG CGGTTCTTAG CTTGCTCTC CTACCACTTC
2451 AGTACCTTCT CTCTTCCCT GGCTCTGAC ATCATTGCA ACAGGAACAT
2501 GGGGAAGCCA GCCCAGCCTG CCCTGAGCG GAGGAGCTG GAAGCACTCT

```

```
2551 TCCTCCCTTA TGACCTGAAG CGGCTGGAGA TGTATTACG GAATATGGTG
2601 GACTATCACC TCATCATGGA CATGATCCCG GCCATCTCTC GCATCTATTT
2651 CCTGAACGAG CTGGGGGACC TGGCCCTGTC TGGGGCTCAG TCGGCTCTTC
2701 TCTTGGGGAT TGGCCTGCAG CATAAGTCTG TGCACGACCT GGAAGAGGAG
2751 ATTGAGCTGC CTCTGGGCGA GTTGAATGGA CTTTGAACCC GGATCATCCG
2801 CAAAGTTGTG AAGCTATTTA ATGAAGTTCA GGAAGAGGCC ATTGAGGAGC
2851 AGATGGTGGC AGCGAAGGAT GTGGTCATGG AGCCACGAT GAAGACCCCTC
2901 AGTGACGACC TAGATGAAGC AGCAAGGAA TTTCAGGAGA AACACAAGAA
2951 GGAAGTAGGG AAGCTGAAGA GCATGACCT CTCTGAATAC ATATCCCTG
3001 GGGACGATGA AGAGTGAAT GAAGTTTGA ACAAGCTGG GCCGACGCC
3051 TCGATCATCA CCTGAAAAG TGACAGAAA AGGAAGTTAG AGGCCAAMCA
3101 AGAACCCAAA CAGAGCAAGA AGTTGAAGAA CAGAGAGACA AAGACAAAA
3151 AAGATATGAA ACTGAAGCCG AAGAAATGT GAGAGAAAC TCGGCAATCT
3201 GTGTTGATC ATGGGAAGAT ACTCTACTA ACTGAACCTT CTCTGGCTGG
3251 ACTGTTAAAA GCAACGAGAG GCCCGGCCAC ACTGGAAGC TGGCGCGAA
3301 TTCGGCTCTT GGGCTGTGT GTCTGTGAGC TCAACCTGGC TAAAGGCGAA
3351 GTCACTGCGA AATGGGCTCT TTGAACTT GATGGGTGG CACTGGCATC
3401 TCTAGATTG CCACGAGTCT CTCTCTTCTT GCCCAGTCCA GGGCCCTCTT
3451 TTCCTATAAG TTCATATTTT GCTTTGAGCC AGCTTTTAG TCTATTCCC
3501 ACACATGTGG AAGCCACGTT GCCTCTCGAC CGCCTGAGGC CTTAAGTAC
3551 ATCCCTTTCT GTGTGTGCC AGGAGCTGCT TCGTGGGCGC CTGGGTCTCT
3601 CTTTGTGGAC TTGTACCTGG AGCAGAGGGA ACTCCAGTCC GTCCGGCAT
3651 CCATGGCAGC CCGCGGTTAG GTGCGCCAGG GTTGCTGAT GTTGCTTGT
3701 GCTGTTCCAC TCTTGGCTCC AGCAGACCCA CTGTCCGAGA AAGCTGTAT
3751 CCTGTAGTTT ATGTAGAAAT CCACATCTGC GTCTCGAAGA CTTGTTTCA
3801 CCATTGGAAG AAGATCTTGT GGAAGGCCA CTTTCTGCG AGGGGTGAGG
3851 GGAAGGATAG AGAATCTATT TTTAATAAT AACATTCTAG AATGAAAAAA
3901 AAAAAA AAAA AAAA AAAA AAAA AAAA AAAA AAAA AAAA
3951 AAAAAA AAAA
```

## BLAST Results

No BLAST result

## Medline entries

No Medline entry

## Peptide information for frame 3

ORF from 102 bp to 3176 bp; peptide length: 1025  
Category: similarity to unknown protein  
Classification: unclassified  
Prosite motifs: RGD (966-969)  
ATP\_GTP\_A (284-292)

```
1 MRRKVKONRI RILICNVAR RORSLFYVVG DRKQDVVIL HHMLSKATVK
51 ARPSVLWVK KELGSSHRK KMRDLQKKI KNGTLMKQD DPFLFLAAT
101 NIRYCYNET HKILGNTFGM CVLQDFEALT PNLLARTVET VEGGGLVLIL
151 LRTNLSLKLQ VVTNVDHRSR YRTEAHODVV GRFNERFILS LASCKKCLVI
201 DDOLNILPIS SHVATNEALP PTFPDESIGP SDLELRELKE SLQDTOPGVV
251 LVDCCKTLQD KAVLKFIDG ISEKTLASTV ALTAANGRGR SAALGLAIAJ
301 AVAFGYSNIF VTSFSPDLNH TLFETVFKGF DALQYQHLSD YEIIQSLNFE
351 FWKAVIRVHV FRENROTIOY IHPADAVKLG QALVVIDEA AAIPLPLVKS
401 LUGPYLVFNA STINGYEGTG RSLSLKLIQD LRQGAOSQV STTAENKTTT
451 TARLASARFL NIVSLOESIR YAPGDVEMW LNDLLCLOCL NITRIVGCCP
501 LPEACELYVV NRDTLFCYHK ASEVFLQRLM ALYVASHYKN SPNDLQMLSD
551 APAHHLFCLL PPVPPTONAL PEVLAVIQVC LEGEISROSI LNSLSRGKKA
601 SGDLIPMTVS EQFQDPDFGG LSGGRYVRIA VHPDYQMGY GSRALOLLQM
651 VYECRFPCLL EKVELTPQEI HTVSSSAVSL LEVITPRND LPFLALLME
701 RPAERLDYLG VSYGLTPRLI KFKRAGEFV VYLRTFPNDL TGEHSCIMLK
751 TLTDEDEADQ GGMIAAPMKD FRRFLALLS YQSTFSPSL ALNIIQNRNM
801 GRPAQPALSR EELEALPLPY DLKRLHMYSR NHVDVHLMD HIPAISRIYF
851 LAQLGDLALS AQSALLGT GLGRKSVQOL EKSELEPSQD LAGLPLRIIR
901 KVKLFLNEVO EKAIDEQWVA AKDVVMEPTH KTLSDOLDEA AKFDEKHKK
951 EVGKLKSMOL SEYIIRGODE ENNEVLNKG PNASIISLKS DKRRKLEAKQ
1001 EPKQSRKLKN RETNKKDKM LKRRK
```

## BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZptes3\_6c11, frame 3

TREMBL:CEAF3130.4 gene: "F55A12.8"; *Caenorhabditis elegans* cosmid  
F55A12., N = 1, Score = 2782, P = 1.1e-289

PIR:S55151 probable membrane protein YNL132w - yeast (*Saccharomyces cerevisiae*), N = 2, Score = 2549, P = 3.5e-273

SWISSPROT:YXK1 ACHAN HYPOTHETICAL PROTEIN (FRAGMENT), N = 1, Score = 1013, P = 3.2e-102

SWISSPROT:YDK9 SCHPO HYPOTHETICAL 116.5 KD PROTEIN C20G8.09C IN CHROMOSOME I., N = 1, Score = 2843, P = 3.8e-296

>SWISSPROT:YDK9\_SCHPO HYPOTHETICAL 116.5 KD PROTEIN C20G8.09C IN CHROMOSOME I.  
Length = 1,033

HSPs:

Score = 2843 (426.6 bits), Expect = 3.8e-296, P = 3.8e-296  
Identities = 576/1033 (55%), Positives = 750/1033 (72%)

Query: 1 MHRKKVDNRIRILIENGVAERQSLFVVVGDGKQDVILHMLSKATVKARPSVLNMYK 60  
M +K +D+RI LI+NG +E+QRS FVVVGDR +DQV LH +LS++ V ARP+VLM YK  
Sbjct: 1 MPKALDSRIPTLIRNGCOEQRSFFVVVGDRADQVNLHMLLSQSKVAARPNVLMHYK 60

Query: 61 KEL-GFSHRRKKMRQLQKKIKNGTLMIKQDDPFELFIAATNIRYCYNETHKILGNTFG 119  
K+L GF+SHRKKR +++K+K G + +DPFELF + TNIRYCY E+ KILG T+G  
Sbjct: 61 KDLGFTSHRKKRENKIKKIKNGIRPNSEDPFELFCSITNIRYCYKESEKILGQTYG 120

Query: 120 MCVLQDFEALTPNLLARTVETVEGGGLVILLATNLSKQLYTVTMDSRYRTEAHQDV 179  
M VLQDFEALTPNLLART+ETVEGGG+VV+LL +NSLKQLYT++ND+HSRYRTEAH DV  
Sbjct: 121 MLVLQDFEALTPNLLARTIETVEGGGIVVLLLNKLSKQLYTHSMOHSRYRTEAHSDV 180

Query: 180 VGRFNERFILSLASCKKCLVIDDQNLILPISSHVATHEALPQTPDESIGPSDLEIRELK 239  
RFRNERFILSL +C+ CLVIDD+LN+LPIS ++ALPP +++ + ++EL+  
Sbjct: 181 TARFNERFILSLGNCENCLVIDDELNVLPIGG-KNVKALPTLEEDN--STQNSIRELQ 237

Query: 240 ESLQOTOPGVVLVDCCKTLQDAKAVLKFIKISGKTLRSTVALTAARGRGKSAALGLAIA 299  
ESL + P G LV KTLQDA+AVL F+E I EK+L+ TV+LTA RGRGKSAALGLAIA  
Sbjct: 238 ESLGEDHPAGALGVYKTLQDAKAVLTFVESIVEKSLKGTVSLTAGRGKSAALGLAIA 297

Query: 300 GAVAFGYSNIFVTSPPDNHLTLFETFKGFDALQYQENLDYELIQSLNPEFNKAVIRVN 359  
A+A GYSNIF+TSPSP+NL TLFEF+FKGFDAL Y+EH+DY+IIOS NP ++ A++RVN  
Sbjct: 298 AAIANGYSNIFITSPPENLKLTFETFKGFDALNVEEHVDYDIQSTNPATNNAIVRN 357

Query: 360 VFREHRTIQIYHPADAVKLGQALVVIDEAAAIPLPLVKSLLGPYLVFNASTINGEYGT 419  
+FR+HRTIQIYI P D+ LGQALVVIDEAAAIPLPLV+ L+GPYLVFNASTINGEYGT  
Sbjct: 358 IFRDHRTIQIYISPEDSNVQALVVIDEAAAIPLPLVKKLIGPYLVFNASTINGEYGT 417

Query: 420 GRSLSLKLQQLQOASQSGVSTTAENKTTTTLASARTLHEVSLQESIRYAPGDAVEK 479  
GRSLSLKL+QQLR+OS S + NK+ + + + S RTL E+SL E IRYA GD +E  
Sbjct: 418 GRSLSLKLQQLREQSRI--YSGSGNKSQSQSHI-SGRTLKEISLDEPIRYANGDRIEL 474

Query: 480 WLNDLLCLDCLN-ITRIVS-GCLPEACELYYVNRDTLFCYHKASEVFLQRLMALYVASH 537  
WLN LKCLD + ++R+ + G P P C LY V+RDTLF YH SE FLQR+M+LYVASH  
Sbjct: 475 WLNKLLCLDAASVSRNATOGFPHPFSECLYRVSRTLFSYHPISEAFLQRMMSLYVASH 534

Query: 538 YKNSPNDLQMSDAPAHKLFCLLPVPTONALPEVLAVIQVCEGIESRQSLNLSLRG 597  
YKNSPNDLQ++SDAPAH LF LLPPV LP+ + VIO+ LEG ISR+SI+NSLSRG  
Sbjct: 535 YKNSPNDLQMSDAPAHQLFVLLPPVDLKNPLPDPICVQLALEGISRESINNSLRG 594

Query: 598 KASGDLIPWTVSEOFQDPDGGLSGGRVRIAVHPDYQMGYGSRALQLQNYEGREFP 657  
++A GDLPW +S+OFQD +F L G R+VRIAV P++ HGYG+RA+QLL Y+EG+P  
Sbjct: 595 QRAGDLPWLSIQOQFQDENFAALGGARIVRIASPEHKMGYGTAMQLLHEYFEGKFI 654

Query: 658 CLEKXVLETPOEINTVSSEAV---SLLEEVITPR--KOLPPLLLKLERPAERLDYLGVS 712  
E+ + + E+ +L E I R K +PPLLLK+E E L Y+GVS  
Sbjct: 655 SASSEFKAVKHSKRIQDEEIENTALQTEKIHVRDAMTPPLLLKLELOPELHYGVGS 714

Query: 713 YGLTPRLLLKPKWRAGFVPVYLRQTPNLTGESHCIMLKTLTDEDEADQGWLAAPKDFR 772  
YGLTP L KWK R G+ P+YLRQTP NDLTGEH+C+ML+ L D ML AF ++F  
Sbjct: 715 YGLTPSLQKPKREGYCYLYLRQTPNDLTGECTVMLRVLGRDSE---WLGAFQNFY 770

Query: 773 RRFLLALLSYOFSTFSPSLALNLIQNRNNGKP---AQPALSREELEALFLPYDLKRLNY 828  
RRFL+LL YOF F+ AL+++ N G + L+ EE+ +F YDLKRL E Y  
Sbjct: 771 RRFLLSLGYQREFAAITALSVDACNNGTEYVVNSTSKLTNEEINNPFESYDLKRLSEY 830

Query:	829	SRNNVYHIDNDPQIPAIRIFVYVWQLQDQ-LALSAQSGALLGIGIKHSVQOLEIEFLP	887
		5 N+DTH+I+D+P+ +F + + D L S O +LL+ +G+G+K+ +L+ +E+ L+ E	
Subject:	831	SNMLLDYHVIVDLPKLKLVYFGSGKFFVQSGVQVLLAGLQVYITDITLKEFLP	890
Query:	888	SQGLWLFNRIIRKRVKLVFNQQAIEQGVAAKDVYNE-----PTMKT+SDDDLE	930
		3 Q+L+ + +K+K+ +D+K IEE+ + K P +L+ +L+ +	
Subject:	919	SHMLNKLKSKINCKNICIDIEEELGSKSTESSKLPPTTFVQGLEEQLQ	955
Query:	940	AAEFQ-EHNKKEGVGRKLSMSEY+IGCDEE+H K A N K A	998
		5 A E D E M E M K A N K A	
Subject:	951	GADEAMLAIREQRELINIAOLEKYAIRGEEDN-----KAEN-QIQTKNGARGVSI	1004
Query:	999	KOEPKQSKLL-KNRETKNKDKLKKKK 1025	
		K E +L+ +K+K K K K K	
Subject:	1005	KOEKKRKNLSQASQETKPKPSKFRKFL 1033	

Pedant information for DKFZphtes3 6c11, frame 3

Report for DKFZphtes3 6c11.3

```
[LENGTH] 1025
[MM] 115704.57
[PI] 8.30
[MOMOL] PIR:555151 probable membrane protein YNL132w - yeast (Saccharomyces cerevisiae)
O.0
[FUNCTION] 10.99 other signal-transduction activities [S. cerevisiae, YNL132w] 0.0
[FUNCTION] cAMP-gate function prediction [H. influenzae, H1254] 2e-05
[PROSITE]
[PROSITE] ATP_GTP_A 1
[PROSITE] RGD 1
[EW] Alpha Beta
[EW] LOW_COMPLEXITY 11.80 %
```

```

SEQ      MHHKKVDNRNRIILITENGVAERQSLFVYVGGRGQGVVLIHHLSKATVKARPSVLMCKV
PRD      ccccccccchhhhhhhccccccccccccccccccccceehhhhhhhhhhhccccceehhhh

SEQ      KELGFSFSHAKKRMRLQKIKNGTLNIRQDDPFELFIATNIRYCYVNIKGLTFGH
PRD      nhhhcccchhhhhhhhhhhhhhhhhhhhhcccccccccccccccccccccccccccccccc

SEQ      CVLDDLPFLPNLLARTVTVEGGGLVVILLRTNHSKQLYTVMDVHSYRTAEHQDVV
PRD      eehhhhhcccccccchhhhhhhhhhhhhccccccccccccchhhhhhhhhhhhhhhhhhhhhhhhh

SEQ      GRNFRNFIILSASCKKVLIDQQLNLPISSHVATNEALPPTOPESLGPISDLEALKE
PRD      hhhhhhhhhhhcccccccccccccccccccccccccccccccccccccccccccccccccccc

SEQ      SLQDTQPGVVLQCCCKTLDAQAVLKFIEGISEKTRTSVLTAAARGCSAALGLAIG
PRD      hccccccccceeeehhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhccccccccchhhhhhhhh

SEQ      AVAGYSNIPVTSPPDNLTFLFIFYFKGDAIYQENLOYEITQSLNPFNKAVIRVNW
PRD      xxx
SEQ      ccccccccccccccccccccchhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhccccccccceehh

SEQ      FRHRTQIYIHPADAPVGLQAEVLVDEAAJLPLPHVLSLGLPVYFVMASTINGEYGT
PRD      hhhhhhhhhccccccccccccccccccccceehhhhhhhcccchhhhhhhcccccccccccccccc

SEQ      RSLSLKIQGLQGSQSGVSTAEHTKTLTALASARTLHVESLQSIIRYAPGDAVEKH
PRD      xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx
SEQ      cchhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhccccccccccccchhhhhhhhhhhhhhhccccchhhh

SEQ      LDNLCLDCLNITRIVSGCPLPEACELVYVNRDITFCYKASEVFLQRLNIVYASVYHH
PRD      xxxxxxxxxxxxxxxx
SEQ      hhhhhhhccccccccccccccccccccccccccccccccccccccccchhhhhhhhhhhhhhhcccc

SEQ      SPNLDLQSDPAHHFLPLPVPTQNALPEVLAVICQVCGEHSISQSLNLSISGKKA
PRD      cccccccccccccccccccccccccccccccccccccchhhhhhhhhhhhhccccchhhhhhhcccccc

SEQ      SGBLIPMTVYSEQFQDPQFGLSGSRVVRIRVHPDQYGGYGSRAQLQQLQYEGNPFCLC
PRD      cccccchhhhhhhhhhhhhhhhhhhhhhhccccccccccccccccccccchhhhhhhhhhhhhhhcccc

SEQ      EKVLTEPQZHTVYSSEAVSLVITPRKDLPLLLKNEAPRHHLLQYLGVSGLCPPLH

```

WO 01/12659

PCT/IB00/01496

```
SEG .....XXXXXXXXX.....
PRD hhhhhcccccehhhhhhhhhhhhccccccccccccccccccccccccccccccccchhh
SEQ KFWKRGFVVPVYLAQT PWDLTGKSCIMLWTLTDEDEADQCGWLAAPKDFRRFLALLS
SEG .....
PRD hhhhhcccccecccccccccceccccccccccccccccccccccccccccccccchhh
SEQ YQFSTFSPLALWIIQNRHMGKPAQPALSRELEALFLPYDLKRLEMYSRNMPVDYHLINO
SEG .....
PRD hhhhhcccccehhhhhhhhhhhhccccccccccccccccccccccccccccccccchhh
SEQ MIPATSRIFYFLNQLGDLALSAQSALLLGIGLQHKSVQLEKETELPSQQLMGLFWRIIR
SEG .....XXXXXXXXXXXXXXXXX.....
PRD hhhhhhhhhhhhhccccccccccccccccccccccccccccccccccccccccccccccccccccchhh
SEQ KVVKLFNEVQEKAIIEQMYAAKDVVMEPTMTLSDDLDEAAKEFOEKHKKEVGLEKSMDL
SEG .....
PRD hhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhcc
SEQ SEYIIRGDDEEMNEVLNKGPNASIIISLKSOKRRLKLEAKQEPKQSKKLKRETNKKDKK
SEG .....XXXXXXXXXXXXXXXXX.....
PRD cccccccccccccccccccccccccccccccccccccccccccccccccccccccccchhh
SEQ LKRKK
SEG XXXXX
PRD hhccc
```

Prosite for DKFZphtes3\_6c11.3

```
PS00016 966->969 RGD PDOC00016
PS00017 284->292 ATP_GTP_A PDOC00017
```

(No Pfam data available for DKFZphtes3\_6c11.3)

DKFZphtes3\_6d16

group: testes derived

DKFZphtes3\_6d16 encodes a novel 695 amino acid protein nearly identical to a sequence from human PAC Clone WUGSC:H\_DJ1185107.2.

The cDNA is different to the proposed gene model: it contains additional exons.  
No informative BLAST results: No predictive prosite, pfam or SCOP motif.

The new protein can find application in studying the expression profile of testis-specific genes.

WUGSC:H\_DJ1185107.2, differences to genmodel

differences to genmodel of WUGSC:H\_DJ1185107.2 two exons skippt.

Sequenced by BMFZ

Locus: /map="7q11.23-q21"

Insert length: 4572 bp

Poly A stretch at pos. 4540, polyadenylation signal at pos. 4520

```
1 GGGGGGGCTA GCTTCGGAGT CTCCCGCGCG CACCTCAGCC GCTCTCAGC
51 GGGGGGGCGC TCGCTCCATC GCCTAAAATG ACCAATGTGT GATTTCAGTG
101 GAATAAATGG CGTCCAAAGT CACAGATGCT ATAGTCTGGT ATCAAAAGAA
151 GATTGGAGCA TATGATCAAG AAATATGGGA AAAATCTGTT GAACAGAGAG
201 AAATCARGGG CTAGAGAT AGACCAAGA AACACGACA TGTGAACCA
251 GACCTCATAG ATGTTGATCT TGTAAAGGGG TCTGCATTGG CAAGGCCAAA
301 GCCTGAAAGT CTTTGGACTT CTCTGACCAG AAAGGGAATT GTTCGAGTTG
351 TATTTTTCCT CTTTTCCTTC CGGTGGTGGT TACAAGTAC ATCAAAAGTC
401 ATCTTTTCTT GCGTCTCTGT CTTTATCTCT CTTCAGTTTG CTGAATAGT
451 ATTTATCTGC TCCACTTCTA GCGCCACAGC CATACTCTG ACAGAGGTGA
501 TTGGGCGCAT ATGGCTGATG CTGCTCCTGG GAACGTGGA TGGCAGATT
551 GTTTCACACA GAACACCCAA ACCTCTCTTA AGTACAGGGG GTAAAGAGAG
601 AAGCAAAATTA ACAAAGGAG CCCATTTGGA AGTACATAGG GAAGGAGATG
651 GTTCTAGTAC CACAGATAAC ACACAAGGGG GACCACTTCA GAACCCGGT
701 ACAAGCACCT CTCACAGCGT TGGCACTGTC TTCAGAGATC TCTGGCATGC
751 TCGTTTCTTT TTATCAGGAT CAAAGAAAGC AAAGAATTCA ATTGATAAAT
801 CACTGAAGAC TGACAATGGT TATGTATGCC TTGATGGGAA GAAGACTGTT
851 AAAGCGGGTG AAGATGGAAT ACAAAACCAT GAACCTCAGT GTGAAACTAT
901 TCGACCAGAA GAGACAGCCT GGAACACAGG AACACTGAGG AATGGTCCTA
951 GCAAGATATC CCAAGAGACA ATAAACAATG TCTCTGATGA AGTCTCCAGT
1001 GAGAGAGGTC CTGAACAGG ATACTAATTA COTGCTCATG TGGCAGGAG
1051 TTCTGAAGGT GTTCTTGGGA ATAGAAAGTC ACACCATTAT AAGAAACATT
1101 ACCCTAATGA GGACGCCCTT AAATCGGGTA CTAGTTGAG CTCTCGCTGT
1151 TCAAGTTCCA GACAGGATTC TGAGAGTGCA AGGCCAGAA CTGAACACAG
1201 AGATGTGTGA TGGGAAGACT TTTTCAATGC TGCAGAAAGC CATTCACTTT
1251 GTACCAGTGA CACAGATGTG GAAATCATC AGATTATCC ATGTGTGAAA
1301 AAAGAATATA GAGATGACCC TTTTCATCAG AGTCATTGCG CCTGGCTCCA
1351 TAGTTCCTCC CAGGATTTAG AAAAATAAG TGCTATAGTA TGGGAAGTGA
1401 ATGATGTGTA GAAGCAGAC ATGCTCTGAC TTGAATCAG TGGATGATA
1451 ATGAACAGAG TGAACAGCCA TATACCAGGA ATAGGATACC AGATTTTGG
1501 AATGTCAGTC TCTCTATAC TGGGTTTAA TCCATTGTTT TTCCGACTTT
1551 CTCAGCTAC AGACTTGGA CAATCTCAG CACATTCTGC TTGAGAACTT
1601 TATGTGATTG CATTGGTCT TAATGAAGAT GTCATAGTTC TTCTATGTT
1651 TATAATAAGT TTTGTGTTTC GGTGTCTCT TGTGTGGATT TTCTTTTTTT
1701 TGCTCTGTGT AGCAGAAAGA ACTTATAAAC AGCGATTACT TTTTGCAAAA
1751 CTCTTTGGAC ATTTAATGAT TGCAGGAGG GCTCGAAAT CTGAGGTTCC
1801 TCATTTCGGG TGAAGAAAG TACAGATAT AAAAATGTG CTATCTCTCC
1851 GTTCTATCTT TAAGCGTCTG GGTCTCAGC GATCAGTTGA TGTAAAGTCT
1901 TCATCTGCTT TCTTATTGAC TATCTCAGT GTATTATCT GTTGTCCTCA
1951 GATAAACCTC TACTTGAAA TGGAGAAAA ACCTACAAA AAGGAGGAGC
2001 TACACTACTG GAATTAATCT TTAATACTGG CTACTAACT GCTAAGGAG
2051 TTGACAGCTC CTTTATGATT ATATGGGCTT ACAATGAAT CGCTGCTTTA
2101 TAACATCACC CAGGTGTGTA TCTGTGAGC TGTTTCTGTT GTTATCAGTG
2151 ACTTGCTTGG ATTTAATTTA AAGCTATGGA AGATTATGTC ATGACAATTC
2201 AAAGAAAGCA AGATCTAGCC TCTTTTCCAG AATAAGATGA CTGACTAAGC
2251 TCTCTGAAGC CTTTCTCAGT ATTTCTTGCT TCAGGAGTCT CAGCTAGGGA
2301 GTTGAAGTGT TTACATCAGA CTGTCTTGT CAATTCTTAT ATTTATTTTA
2351 CTGGTCTACT TTTTITTTTACA TTTATTTTAG TCTTATATT TTTATTTTTA
2401 AGCATTTGAT TACTTATGTT TTAAGAGGCT GATGAAGTCT ATATCCAGAT
2451 ACTTGAGATC TGGGTAAATG GTCAATATA ATTGCAAAA TAACAAATTC
2501 TGAATAATGA AGCCATTGCT CAGCACCGTT TCTCCATCAA TGCCGTGAAC
2551 TTGCTTACTT TGAGAAAAAA TTCTTTAACT TTGGAATATT GCATTGAAC
2601 CAGCTATACA CATAAAAAAT TTCTTTTGGT AAATCAAGAT CCACTCAGGG
```

```
2651 TTCTCTTGA ATTATTTGG AACATGCCA GGATCCAAC TGATTAAGTT
2701 ACAGTTTAA GACCTTCAC TATTAATATA TACGGTATA TATACAGGT
2751 CACCAAGTGC TCTTTGATG TAAACTTGT AATGAGCAA TAATTCTAAA
2801 TGGTTACCAT ACTGAAGAT ATTTTGATA AAATTAACCTA GTAATCTTG
2851 TATTTATTGG AACACCTGGG CTGTTTGAC AGCTCCAAT GTGATGCTC
2901 AAAATGCGCA CTTTAAAAA TTGTTACTTT TAATGGCTAT CTTTATATGG
2951 GATCTGTTAT AGTATACAG GGCATGATAT GGTATCCTTT TGAGTGAGGT
3001 ATATACATCT CTCACAAGT AAGTGCTAC TGATATTACT AAAGTACATT
3051 ATGTTTACTC AAGTAAATAA TTTTCTCCC ATGGTACACT CTAGTGAGG
3101 CTATTCATAC CACACTGAAA TGACCACTG AAGATTAAGG CTACAACCA
3151 ATAAATATTT TCTCTAATTG CTAGTTGTA AACTCTATCC AAATTTTCAG
3201 AAAAGACAGC TTCAGCTTGC AAATCTATC CTCTAAACTT ATCTGGTGCA
3251 TTCTCCCCAC CCCACCCCCA TTATATAAGG GCTATTTTAG ATGCTTTTAA
3301 CTTCCCAAC AAATAATTTG CCAGATGTC AATGAGAACT TATCATGCTG
3351 CTGTGTTAGG TAACTGGGC AAATAGATA GTTCTTACA TTGGGCTTG
3401 ATTTTAAGTT GTTATATTTG TACAATCGAG TATTTAGAA ATTACATGAA
3451 ACATGAACAA GTTTTGCAG TTTTFTTAA ACTGGGCATC TGGTTCTTAA
3501 AAATTTATTT GAAACATCT AGAATTTCT TGCTGCAAG TGTATCATGT
3551 GGAATATCT CATATTTTAA CCATATTTTA AGAATTTTAA GACGATTAAT
3601 TGTAAATAAT TTATTTGATT GTGCGATTC TAATCCCTAA ATCATAATCT
3651 TAAATTCAGG AATGTGTGGA GAACAGAGCC ATGTCTATATC ACTTTGCTCT
3701 TACCATCTCT TTGATCAGC CTCATTCAG CCTCATGTG TAGTATGTTT
3751 TTTCTTCTTA TGAAGAACAA CAGAAAGCAT TTTATTTAT TTGCTATGT
3801 TCAAAATATG TTAATAATGA CCAAGTGCA TTCTGAGTT TTCAAGGAA
3851 TGTAAATCTG GAGCTTTAAG AACATCTTA GTTCTCATG TGAAGACTTA
3901 GCGTTGTGCT GATGTTTTC CTCTCTATG TGTCTAATG TGAGTGTGT
3951 TTGAGAAAT ATGTTTATA AACTTTTCA ATATAAGTGA CATGCTTATA
4001 CAGAACTTAA CATTITGCAC AGAATATATC AAATATATT TGAGAAAAAA
4051 AGTACGGCAT GAGTTCTGTT AGGAATAAAA GATGAACCTA TTGATCTCA
4101 CAAAAAATCT TATTTCAGAA TGAATATATT TTTGAGAAA GTAGCTGAGT
4151 ATACTGTTT AAGAAATGC TTGTTTATG TTGAGGTTAA CTTAGAGTTG
4201 GGAGTGATT TATTAAGTAC AGTATACCTC TCAACAGTT ATAAATAATA
4251 TGTGAATTA TGTAGTGTG GGCAGCAGTA GAATCTAATA AGGAAATGT
4301 CATGTTAAGC AATTTCAGAA CATTACTGA ACTATTCTCA AAGCAGAAA
4351 ATTGACATTG CTGCTTTTAA GAATACCATC AATGTAGAA ATTGAAGAA
4401 ATTGTAAAT ATCACATAAT ATAGAAATG CAGTCAAG AGAATTTGCG
4451 CAGATGTTGT GTGGAACCTG TTGTTTCTT GCCACATGT TTGATTTGA
4501 AAGTTTACA GTAAATTTAA AATAAAGAT TCTGTACTG AAAAAAAA
4551 AAAAAAAA AAAAAAAA AA
```

## BLAST Results

No BLAST result

## Medline entries

No Medline entry

## Peptide information for frame 2

ORF from 107 bp to 2191 bp; peptide length: 695  
Category: known protein  
Classification: unclassified  
Prosite motifs: CYTOCHROME\_C (375-381)

```
1  MASKVTDAIV WYQKIGAYD QQWEKSVED REIKGLRKP KTAHVKPD
51  IDVLRGSA FAKAKPESFV TSLTRKIVR VVFPFFFRW MLQVTSKVIF
101  FWLLVLLQ VAAIVLFCST SSPHSIPLTE VIGPIWMLL LQTVHCQIVS
151  TRTPPLST GKRKKRLAK AAKLEVHREG DGSSTIDTQ EZAVQHQDTS
201  TSHSVCTVR DIRMIAATLS GSKAKRSLD KSTETONGYV SLGCKPVFS
251  GEDGQNHPE QETIRPEET AMWTGLRNG PSKOTQIT NVSDEVSSEE
301  GPETGYSLRR HVDTSEGV L RNRKSHHYK HYPNEDAPRS QTSRSSRCS
351  SRQSEARP ESDTEVME DLLHCAEHS SCTSETDVEN HQINPCVRKE
401  YRDPFHQSH LPMHSHHPG LKLSIAIWE QNCKKADMS VLEISQINI
451  RVNSHIPGIG YQIFGNVSL ILGLTPFVFR LSQATLEQL TAHSASELYV
501  IAFGSNEDVI VLSHVLIIFV VRVSLVWIFF FLICVAERTY KORLLFARLF
551  GHLTSARRAR KSEVPHRLK KVNQIKRML LSTLRNGP QRSVDIVS
601  AFLITISVF ICACIHLVL RMEKPKKE ELTLVNVKL LATKLLKELD
651  SPFRLYGLTM NPLLYNITQV VLSAVSGVI SDLLGNLKL MKIKS
```

## BLASTP hits

WO 01/12659

PCT/IB00/01496

No BLASTP hits available

Alert BLASTP hits for DKFZphtes3\_6d16, frame 2

PTR: S38170 SRP40 protein - yeast (Saccharomyces cerevisiae), M = 1,  
Score = 100, P = 0.00TREMBL:AC004990.1 gene: "WUGSC:H\_DJ1185107.2": Homo sapiens PAC clone  
DJ1185107 from 7q11.23-q21, complete sequence., N = 2, Score = 2693, P  
= 0>TREMBL:AC004990.1 gene: "WUGSC:H\_DJ1185107.2": Homo sapiens PAC clone  
DJ1185107 from 7q11.23-q21, complete sequence.  
Length = 588

HSPs:

Score = 2693 (404.1 bits), Expect = 0.0e+00, Sum P(2) = 0.0e+00  
Identities = 510/515 (99%), Positives = 512/515 (99%)

Query: 35 GLRNKPKKTAHVKPDIDVLRGSAFAKAPESPMTSLTRKGIVRVVFVFFFRWMLQV 94  
Sbjct: 1 GLRNKPKKTAHVKPDIDVLRGSAFAKAPESPMTSLTRKGIVRVVFVFFFRWMLQV 60

Query: 95 TSKVIFFWLLVYLQVAAIVLFCSTSSPHSIPLTVIGPIWMLLLGTVMCQIVSTRTP 154  
Sbjct: 61 TSKVIFFWLLVYLQVAAIVLFCSTSSPHSIPLTVIGPIWMLLLGTVMCQIVSTRTP 120

Query: 155 KPPLSTGGKRRRLRKAHLEVHREGDGSSTTONTQGAQVQNHGTSTSHSVGTVFRDLWH 214  
Sbjct: 121 KPPLSTGGKRRRLRKAHLEVHREGDGSSTTONTQGAQVQNHGTSTSHSVGTVFRDLWH 180

Query: 215 AAFFLSGSKKAKNSIDKSTETONGVVSLOGKRTVKSDEGQIQNHPOCETIRPEETAMNT 274  
Sbjct: 181 AAFFLSGSKKAKNSIDKSTETONGVVSLOGKRTVKSDEGQIQNHPOCETIRPEETAMNT 240

Query: 275 GTLRNGPSKDTORTITNVSEVSSEEGPETGYSLRHHVORTSEGLVRNRKSHHYKKHYPN 334  
Sbjct: 241 GTLRNGPSKDTORTITNVSEVSSEEGPETGYSLRHHVORTSEGLVRNRKSHHYKKHYPN 300

Query: 335 EDAPKSGTSCSSRCSSRRQDESARPESETEDVWEDLLHCAECHSSCTSETDVENHQIN 394  
Sbjct: 301 EDAPKSGTSCSSRCSSRRQDESARPESETEDVWEDLLHCAECHSSCTSETDVENHQIN 360

Query: 395 PCVKKEYRDDPFHQSHLPWLHSSHPGLEKISAIVWEGNDCKKADNSVLEISGMIMNRVNS 454  
Sbjct: 361 PCVKKEYRDDPFHQSHLPWLHSSHPGLEKISAIVWEGNDCKKADNSVLEISGMIMNRVNS 420

Query: 455 HIPGIGYQIFGNAVSLILGLTPFVFRLSQATDLEQLTAHSASELYVIAFGSHEDVIVLSM 514  
Sbjct: 421 HIPGIGYQIFGNAVSLILGLTPFVFRLSQATDLEQLTAHSASELYVIAFGSHEDVIVLSM 480

Query: 515 VIISFVVRVSLWIFFFLCVAERTYKORLLFAKL 549  
Sbjct: 481 VIISFVVRVSLWIFFFLCVAERTYKQINLYLKM 515

Score = 409 (61.4 bits), Expect = 0.0e+00, Sum P(2) = 0.0e+00  
Identities = 92/115 (80%), Positives = 98/115 (85%)

Query: 595 DVIVS---AFLLTISVVF-----CCA-----QINLYLMEKKPNKKEELTLVNNVLK 640  
Sbjct: 474 DVIVLSHVIISFVVRVSLWIFFFLCVAERTYKQINLYLMEKKPNKKEELTLVNNVLK 533

Query: 641 LATKLLKELDSPFRLYGLTNPLLYNITQVILSAVSGVISDLGLFNKLKWKIKS 695  
Sbjct: 534 LATKLLKELDSPFRLYGLTNPLLYNITQVILSAVSGVISDLGLFNKLKWKIKS 588

Pedant information for DKFZphtes3\_6d16, frame 2

Report for DKFZphtes3\_6d16.2

(LENGTH) 695  
(MW) 78466.68  
(pI) 9.30  
(HOMOL) TREMBL:AC004990.1 gene: "WUGSC:H\_DJ1185107.2": Homo sapiens PAC clone DJ1185107  
from 7q11.23-q21, complete sequence. 0.0



WO 01/12659

PCT/IB00/01496

```

(POSITE) CYTOCHROME_C 1
(KW) TRANSMEMBRANE 6
(KW) LOW_COMPLEXITY 5.32 1

SEQ MASKVTDAIVVYQKKIGAYDQIWEKSVEQREIKGLRNKPKKTAHVKPDLDIVDLVRGSA
SEG .....
PRD cccceeehhhhhhcccccchhhhhhhhhhhhhccccccccccccccccccccccccccch
MEM .....

SEQ FAKAKPSPWTSLTRGIVRVVFFFFFWLQVTSKVIFFMLLVLYLQVAAIILFCST
SEG .....
PRD hhhccccccccccccccccccccchhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhheec
MEM .....

SEQ SSPHSIPLTEVIGPIWMLLLGTVHCQIVSTRPKPLSTGGKRRRLRKAHLEVRREG
SEG .....
PRD cccccccccchhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhheec
MEM .....

SEQ DGSSTDTMTQEGAVQNGTSTSHSVGTFRDLWAAFLSGSKAKNSIDKSTETDNGTV
SEG .....
PRD cccccccccccccccccchhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhheec
MEM .....

SEQ SLGKKTYSGEDGIQWHPQETIRPEETAHNTGTLRNGPSKDTORTITNVSDEVSSE
SEG .....
PRD ccccccccccccccccccccccccccccccccccccccccccccccccccccccc
MEM .....

SEQ GPETGYSLRHVORTSEGVLNRKSHHYKHYFWEDAPKSGTSCSSRCSRRQDESARP
SEG .....
PRD cccccccccccccchhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh
MEM .....

SEQ ESETEDVLMEDLLHCAECHSSCTSETOVENHQINPCVKEYRDPFHQSHLFWLHSSHPG
SEG .....
PRD cccccccccchhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh
MEM .....

SEQ LEKISAIWEGNCKKADMSVLEISGHINHRVNSHIGICVQIFGNVSLILGLTPFVFR
SEG .....
PRD cccccccccccccccccchhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh
MEM .....

SEQ LSQATDLEQLTAHSASELYIAFGSNEQIVLSHVIISFVVRVSLVWIFFLLCVAERTY
SEG .....
PRD hhhhhhhhhhhhhccccccccccccccccccccchhhhhhhhhhhhhhhhhhhhhhhhhhh
MEM .....

SEQ KQRLLFALKFQHLTSARRARKSEVPHFLKQVQIKMWLSLRSYLKRGQPRSVQDIVSS
SEG .....
PRD hhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh
MEM .....

SEQ AFLTISVVFICCAQIMLYLKEKKPKKEELTVNNVLRATKLKELDSPFLYGLTM
SEG .....
PRD eeeeeeeeeehhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhheec
MEM .....

SEQ NPLLYNITQVILSAVSGVISDLGFWLKLKIKS
SEG .....
PRD cchhhhhheeeeeccccchhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh
MEM .....

```

Prosite for DKFzptes3\_6d16.2

PS00190 375->381 CYTOCHROME\_C PDOC00169

(No Pfam data available for DKFzptes3\_6d16.2)

WO 01/12659

PCT/IB00/01496

DKFZphtes3\_72k11

group: testes derived  
DKFZphtes3\_72k11 encodes a novel 233 amino acid protein with similarity to S.pombe  
hypothetical repeat-containing protein.

The novel protein contains 5 leucine zippers and a microbodies C-terminal targeting signal (S-  
K-L) signature. This sequence is responsible for transport of proteins from free polysomes  
into the microbodies.

No informative BLAST results; No predictive prosite, pfam or SCOP motifs.

The new protein can find application in studying the expression profile of testis-specific  
genes.

similarity to S.pombe hypothetical repeat-containing protein

complete cDNA, complete cds, 6 EST hits (3 from testis derived  
libraries)

Sequenced by DKFZ

Locus: unknown

Insert length: 1134 bp

Poly A stretch at pos. 1124, polyadenylation signal at pos. 1088

```
1 AACCTTTCAA GTGCCCCCTC CTTTCCTTAA AGTCTTTTAT AGGGGTCCCC
51 TTCTTGCCCA TCTCCATCCT GTGAGTCAGG ACTGAAAGGG CACAGACAGG
101 TCAGTGCAGG CATTGTTGGG GCAAGCCTGC AAGCAGCAT CACTGGGAT
151 CTGACATGAC AATGCGCCG TGCCTCTCT CAGGGCTACA GGACTTACCC
201 CAGTGGGAAG CAGCTAAGCA GGTCTGACCA GCCGACCTGG ACCTGGCCAA
251 GGTCTCTGTC ATCCCTCATG CCCACCCCGC CATTCCGGCT GATAAGGAAG
301 ATGTTTCTCT TCAAGGTGAG CAGATGGATG GGGCTTCTCT GCTTCCGTC
351 CTTGGGGCCA TCTCTTCCCA GTATTGGCCA GAAGAACTA ATGCACAGC
401 TGCAGGAGGA AAAGGCTTTT CGCGAAGAGA TGAATAATTT TCGTGAAGAA
451 ATAGAGGACT TCAGGGAAGA GATGTGACT TTCCGAGCCA AGATCCATGC
501 TTCCCGGGGC CAGATCTCTG GTTTTGGGA AGAGGAGAGA CTTTCTGGG
551 AAGAGGAGAA AACCTCTGG AAAGAGGAAA AATCCTCTG GGAATGGAA
601 AAGTCTTTCA GGAAGGAAGA GAAACTTTC TGGAAAAAGT ACCGACCTTT
651 CTGGAAGGAG GATAAGGCTT TCTGGAAGA GGACATGCC TTATGGGAAA
701 GAGACCGGAA CTTCTCTCAG GAGGCAAGG CCTGTGGGA GGAAGAAAG
751 GCCCTGTGGG TAGAGGAAG AGCCCTCTT GAGGGGAGA AAGCCTGTG
801 GGAAGATAAA ACCTCCCTCT GGGAGGAAGA GAATGCCCTC TGGAGGAAG
851 AGAGGGCCTT CTGGATGGAG AACATGGCC ACCTTGGCGG AGAGCAGATG
901 CTCGAAGATG GGCCTCACAA GGCACACAGA GGGCAGGCT TCTGGCCTT
951 CTCGGAGGCG AGGGCGTAGC CAGCATGAG GTGCAGGCC CTGTGCTCCA
1001 GACTCCCTGT GGTGGGATT CAAGTCCAGG GTGAGCCCAT GTGCTGGAGA
1051 AAATACACAC TCATTGGTCT CTTGCTTTG AAAGATCCAA TAAAGTCTG
1101 AGGCAAGGTT TGGAAACCA ACTTAAJAAA AAAA
```

#### BLAST Results

No BLAST result

#### Medline entries

No Medline entry

#### Peptide information for frame 1

ORF from 268 bp to 966 bp; peptide length: 233  
Category: similarity to known protein  
Prosite motifs: MICROBODIES\_CTER (231-234)  
LEUCINE\_ZIPPER (142-164)  
LEUCINE\_ZIPPER (149-171)  
LEUCINE\_ZIPPER (156-178)  
LEUCINE\_ZIPPER (163-185)  
LEUCINE\_ZIPPER (170-192)  
LEUCINE\_ZIPPER (170-192)

1 MATPPFLIR KMFSPKVSRL MGLACFRSLA ASSPSIRQKK LMHKLQEEKA  
51 FREENKIFRE KIDFREEMW TFRGKIHAFR GOILGFWEER RPFWEEKETF  
101 WKEEKFWEH EKSFEERTK FWKKYRTFWK EDAFVKEEN ALWERDRNLL  
151 QEDKALWEE KALMVEERL LGEKALWED KTSLWEENA LWEEERAFWH  
201 ENNGHVAEGE MLEDGPHNAN RGORLALFSR GRA

**BLASTP hits**

Entry SPCC330.4 from database TREMBLNEW:  
gene: "SPCC330.04c"; product: "hypothetical repeat-containing protein";  
S.pombe chromosome III cosmid c330.  
Score = 149, P = 1.6e-08, identities = 55/187, positives = 88/187

Entry A45973 from database PIR:  
trichohyalin - human  
Score = 147, P = 3.0e-07, identities = 57/194, positives = 94/194

Alert BLASTP hits for DKF2phtes3\_72kl1, frame 1

No Alert BLASTP hits found

Pedant information for DKFZphtes3 72k11, frame 1

## Report for DKFZphtes3\_72k11.1

[LENGTH]	233	
[MW]	28752.65	
[pI]	5.70	
[PROSITE]	LEUCINE_ZIPPER_5	
[PROSITE]	MICROBODIES_CTER	1
[PROSITE]	MYRISTYL	1
[PROSITE]	CK2_PHOSPHO_SITE	3
[PROSITE]	PKC_PHOSPHO_SITE	4
[KW]	All_Alpha	
[KW]	LOW_COMPLEXITY	15.45 %

```

SEQ      MATPFFRLHMFYSFVSRWGLACIRSLAASPSIRKKHMLKQDEKAFREKRIKFE
SEG      ccccccccccccccccccccccccccccccccccccccccccccccccccccccccc
PRD      KIEDFEHMHZTFRGKIHAHGHQIQLGFEEVAFPEEKETFEWKEKSTWCKHESFREKZET
SEG      ccccccccccccccccccccccccccccccccccccccccccccccccccccccccc
PRD      hhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh
SEG      FWKRYTFPEDEKAFWEDEALMERORLLGDEKALWEERKALWEVZELLEGELERAF
SEG      hhhcccccccccccccccccccccccccccccccccccccccccccccccccccccccc
PRD      KTSLEEMNALWEERKAFMEHNGHVGAGQZLEDGPHANRGORLLAFSGRA
SEG      .RRKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKK
PRD      hhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh

```

## Prosite for DKFZphtesJ\_72k11.1

PS00005	35-17	PKC_PHOSPHO_SITE	PDOC00005
PS00005	35-38	PKC_PHOSPHO_SITE	PDOC00005
PS00005	71-74	PKC_PHOSPHO_SITE	PDOC00005
PS00005	113-116	PKC_PHOSPHO_SITE	PDOC00005
PS00006	106-110	CK2_PHOSPHO_SITE	PDOC00006
PS00006	113-116	CK2_PHOSPHO_SITE	PDOC00006
PS00007	182-187	CK2_PHOSPHO_SITE	PDOC00006
PS00008	81-87	MYRISTYL	PDOC00008
PS00342	231-234	MICROBODIES_CTER	PDOC00029
PS00029	142-164	LEUCINE_ZIPPER	PDOC00029
PS00029	149-151	LEUCINE_ZIPPER	PDOC00029
PS00029	152-178	LEUCINE_ZIPPER	PDOC00029
PS00029	163-185	LEUCINE_ZIPPER	PDOC00029
PS00029	167-192	LEUCINE_ZIPPER	PDOC00029

(No Pfam data available for DKFZphtes3\_72k11.1)

DKFZphtes3\_72k15

group: cell structure and motility

DKFZphtes3\_72k15 encodes a novel 188 amino acid protein with strong similarity to Rattus norvegicus actin-filament binding protein Frabin.

FGD1-related F-actin-binding protein (Frabin/FGD1) is a novel F-actin-binding protein. The gene locus fgd1 seems to be responsible for facio-genital dysplasia or Aarskog-Scott syndrome. Frabin binds F-actin and shows F-actin-cross-linking activity. Overexpression of frabin in Swiss 3T3 cells and COS7 cells induces cell shape change and c-Jun N-terminal kinase activation, as described for FGD1. Because FGD1 has been shown to serve as a GTP/GDP exchange protein for Cdc42 small G protein, it is likely that frabin is a direct linker between Cdc42 and the actin cytoskeleton. Cdc42p is an esin yeast, Cdc42p transduces signals to the actin cytoskeleton to initiate and maintain polarized growth and to mitogen-activated protein morphogenesis. In mammalian cells, Cdc42p regulates a variety of actin-dependent events and induces the JNK/SAPK protein kinase cascade, which leads to the activation of transcription factors within the nucleus. The novel protein seems to be the human orthologue of rat frabin.

The new protein can find application in modulating of cell structure and motility as well as modulation of the JNK/SAPK pathway.

strong similarity to actin-filament binding protein Frabin

2 EST hits

Sequenced by DKFZ

Locus: unknown

Insert length: 1845 bp

Poly A stretch at pos. 1835, polyadenylation signal at pos. 1816

```
1 GTGATGGAGA GTGCTGTTAT GATAGATGAA TCTAGGAAAG CCTCTTTGGA
51 GATGTGATAC CTGAACAGAA CCCCAGATGA TAAGAAGAAA TACCAGTGT
101 TTAGGAGAGA TTGTCTTAAG CAGAGAAGAC CAGCTGCAAA GACCCCAAGA
151 CACATACACT TCGTTATTAA GAATGGCAGC AGCAGAGACT ATGGCAAGAA
201 CACAGTGAAT TTTCCCTTGA GTGTGTGAGG AAGCCCTCAG AGTTTGTGAC
251 TGACTTTGAG AGGTTCTAGT GGAGGGGATC AGAGTGGAAA CAAAGAGACC
301 AGTTAAAGAG GTATGGCAGC ATGAATAAAA AGCTTTTGAG AGTATTCTAT
351 ATGCTCTTCA AATAAAAGAC TCTTTGCTTC ATAATTGCTT CATATAAGTA
401 GGACTGGCTA CACTGTACTA TTTAAAAATG TTAAGAAACA TCAATAAGTA
451 AAAATGTTAG GAAGAGATGA TAAATACGTA AGTATTATAT CTAACATAAGT
501 CTTTACTATG TAGTCACATT ATTAACAGCT GCAAGGATCA AGAAGATTTA
551 AGCGTTGAAA AATAAATAAA TAAGTTATAA ATAAATAAAA CAGCCCAAGG
601 AAATGTTCCA GTCCCATAG GTAGACTCGG GGTCTCTTTC TTTATTTAAA
651 TCTTTATTTA AATGTGGATA GCATCCCAAG AGACTTGGGT CTACACTAAG
701 AATATTCAAA TCCATGTTTC TGAAAGCATC AGAGATAGAA AAAAAGTA
751 GCGAATATCC CTTTTCAACT GGAATAAACT TGTCTTAATT CTAGACTTT
801 TCCATACCAA TOTTTCATG CTTCCTTTGT ATTTTATCTT TTAGCTCAAT
851 ATCAAAATTAT AGTGATTGTA AGAAAGAGTC TGCTGTGAAC CTAAATGCTC
901 CTAGAACCCC AGCAGGCGCT GGAATGACAA CCACACTCTCA ACAAAGCTC
951 CTCTCCAGCC ACTTGCCACA GAGGCGGAGA AATCATACAG ATAAGACTCA
1001 GGGTGCCAGC ACTTGTGTGG CCAACGGTGT AATGGCAGCA CAAAACGAGA
1051 TGGAAATGTA GGAGGAGAAA GCTGCGACTC TTAGCTCAGA TACTCTTATT
1101 CAGCTCTGTC AACCTTCTCT GATATACGAC ATAGTGAATG GAGAAGAGA
1151 TGAAACTGCC ACAGCTCCTG CATCACCCAC AACAGATAGC TGTGATGGAA
1201 ATGCTTCTGA CAGTAGCTAC AGGACTCCAG GCATAGGCCC AGTGCTCCGC
1251 CTAGAGAGAA GAGGGGCGAG AACAGAAACC AGGTTCAAG AGAGGGAGAA
1301 TGGGAAAGCC CCTCTGGAAC TGGAGCAGCT GGACAGGCAC CATGAGATGA
1351 AGGTAGAGCA TGAGACTAGC TCATGAGCAG GGAAACCCCT GCCTATTGCA
1401 TTGTTGCTCT AAAACTCTTT ATTTATTGCA CCCCAGAAAT GTATGAATCA
1451 GATCACCCAC ACTGGCAGTT AATGCTATTT CAGCTCTGG CTCTGATTA
1501 GCATTTCCCC TATGCTCTAA GCAGATATTT CACTTTTCTT TTTCATGTAG
1551 TTCTCTGTAA TATCTCTGTT GTAATTTTCA GAGTCAGAAC AGTGTGGAAA
1601 CTTTAATATA GGAATCCAC AAATGTATTT TTTTACATA GAAGAGAAAT
1651 GTTCTCTGTT GCTCTAGATG TTGGGCTGT ATCCCTATA CTTAGGGGCC
1701 AAGCAAGAGC AAATGTATTA ATCTTTGTTG TTCAGAGTT TCTAATAGAA
1751 TAAATAGGCC TATAAGATGA ACTTGGCACT AGTAAATGTT ACTTTTAGAG
1801 ACATGAATAT GGAAGTATTA AATTATTCAA CGATAAAAAA AAAAA
```

BLAST Results

No BLAST result

## Medline entries

98334590:  
Frabin, a novel FGD1-related actin filament-binding protein capable of  
changing cell shape  
and activating c-Jun N-terminal kinase.

## Peptide information for frame 3

ORF from 810 bp to 1373 bp; peptide length: 188  
Category: similarity to known protein  
Classification: Cell structure/motility  
1 MFSCFLCLLS FSSLSNYSDEL KRESAVNLMA PRTPGRHGLT TTPQKLLSQ  
51 HLPQROGNDT DKTOGAQTCV AHGVNAAQHQ MCEEEKAAAT LSSDTSIQAS  
101 EPLLDTHTVN GERDETATAP ASPTTDSQDG NASDSSYRTF GIGPVLPLEE  
151 RGAETETKVO ERENGESPLE LQLOQHHEM KVENETSS

## BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZpHtes3\_72k15, frame 3

TREMBL:AF038388.1 product: "actin-filament binding protein Frabin";  
Rattus norvegicus actin-filament binding protein Frabin mRNA, complete  
cds.. N = 1, Score = 428, P = 1.8e-39

>TREMBL:AF038388.1 product: "actin-filament binding protein Frabin"; Rattus  
norvegicus actin-filament binding protein Frabin mRNA, complete cds.  
Length = 766

## HSPs:

Score = 428 (64.2 bits), Expect = 1.8e-39, P = 1.8e-39  
Identities = 90/174 (51%), Positives = 115/174 (66%)

Query: 12 SLSLSNYSDELKRESAVNLMA PRTPGRHGLT TTPQKLLSQ HLPQROGNDT DKTOGAQTCV 71  
S L S L S Y T D V Q R D S T H N L I P O T P R H G L T T T P Q K L P S H K S P K O E K D S D Q N Q G H G C L A 90  
Sbjct: 31 SVLSYTDVQKRDSTHNLIPOTPRHGLTSTTPQKLP SHKSPKOEKDSQNGQGHGCLA 90  
Query: 72 NGVAAQHQMECEKEKAATLSSDTSIQASEPLDTHVNGERDETATAPASPTTDSQDGW 131  
NGV AAQHQMECE EK A L S Y T O + D H + NG R + E T T A S T + S D N  
Sbjct: 91 NGVAAQHQMECETEKAAALS PETOTQTAAASPDHVLNGVRNETTDSASSVTNSHDEN 150  
Query: 132 ASDSSYRTPGIGPVLPLEERGAETETKVOERENGESPLELQLOQHHEMKVEHE 185  
A D S S R T G L P + E E + + Q E R E N G S L L Q Q H H E + K + E  
Sbjct: 151 ACDSRCTQTDTLGLPSKEGEPVIEALQERENGLSTEGLNPLQQHHEVKETNE 204

## Pedant information for DKFZpHtes3\_72k15, frame 3

## Report for DKFZpHtes3\_72k15.3

[LENGTH] 188  
[MW] 20388.32  
[PI] 4.62  
[HOMOL] TREMBL:AF038388.1 product: "actin-filament binding protein Frabin"; Rattus  
norvegicus actin-filament binding protein Frabin mRNA, complete cds. 2e-38  
[KW] All Alpha  
[KW] SIGNAL PEPTIDE 16  
[KW] LOW\_COMPLEXITY 12.77 %

SEQ MFSCFLCLLSFSSLSNYSDELKRESAVNLMA PRTPGRHGLT TTPQKLLSQ HLPQROGNDT  
SEG .XX  
PRD cccchhhhhcc  
SEQ DKTOGAQTCVANGVNAAQHQMECEKEKAATLSSDTSIQASEPLDTHVNGERDETATAP

**PCT/IB00/01496**

(No Prosite data available for DKFZphtes3\_72k15.3)  
(No Pfam data available for DKFZphtes3\_72k15.3)

DKF2phtes3\_72p16  
-----

group: intracellular transport and trafficking

DKF2phtes3\_72p16 encodes a novel 796 amino acid protein with very strong similarity to Mus musculus maternal-embryonic 3 (Mem3) gene.

Mem3 was isolated from a partial subtraction library of mouse unfertilized eggs and preimplantation embryos. Its transcript is abundant in the unfertilized egg and also actively transcribed from the newly formed zygotic genome. As Mem3, the novel protein is similar to yeast VPS (vacuolar protein sorting) 35. The null allele of VPS35 results in yeast in a differential defect in the sorting of vacuolar carboxypeptidase Y (CPY), proteinase A (Pca), proteinase B (Prb), and alkaline phosphatase (ALP).

The new protein can find application in modulation the sorting of proteins into different compartments.

strong similarity to mouse MEM3 and yeast VPS35

Sequenced by DKF1

Locus: /map="16p13.3"

Insert length: 2707 bp

Poly A stretch at pos. 2697, no polyadenylation signal found

```
1 CTACGCGCGG GCGGGTGGT GCTTGCTGCA GGCTCTGGGG AGTCGCATG
51 CCTACAACAC AGCAGTCCCC TCAGGATGAG CAGGAAAAGC TCTTGGATGA
101 AGCCATACAG GCTGTGAAGG TCCAGTCATT CCAAATGAAG AGATGCCCTGG
151 ACAAACACAA GCTTATGGAT TCTCTAAAC ATGCTCTAA TATGCTGGT
201 GAACCTCGGA TTCTATGTT ATCACCAGAG AGTACTATG ACCTTTATAT
251 GGCATTCTCT GATGAACGTC ACTACTTGGG GGCTACCTG ACAGATGAGT
301 TTGCTAAAGG AAGGAAAGTG GCAGATCTCT ACGAATCTGT ACAGATGCT
351 GGAACATTTA TCCCAAGGCT TTACCTTTTG ATCAGATTG GAGTGTATTA
401 TGTCAGTCA TTTCTCTAGT CCAGGAAGGA TATTTGAAA GATTTGTAG
451 AAATGTGCGG TGGTGTGCAA CATCCCTTGA GGGTCTGTT TCTTGGAAAT
501 TACCTTCTTC AGTGTACCG AGAATATCTTA CCGATGAAG GAGAGCCAAC
551 AGATGAAGAA ACAACTGGTG ACATCAAGTA TTCAATGAT TTGTACTTGC
601 TCACTTTTGC AGAATGAAC AAGCTCTGGG TGCGAATGCA GCATCAGGGA
651 CATAGCCGAG ATAGAGAAA AAGAGAACGA GAAGACACAG AACTGAGAAT
701 TTTAGTGGGA ACAAAATTGG TGCGCTCAG TCAGTTGGA GGTGTAAATG
751 TCGACGTTTA CAACAGAT TTCTTCTGTC GCATATTGGA CCAACTGTGA
801 AACTGTAGGG ATGCTTTGGC TCAAGAAAT CTATGAGAT GTATTATTCA
851 GOTTTCCTCT GATGAATTTC ACCTCCAGAC TTGGAATGCT TTCTCTGGG
901 CTTGTGCTGA GTTACACCG AATGTAAATG TGAAGACAT AATCATGCT
951 TTAATTGATA GATTAGCTTT ATTTGCTCAC GGTGAAGAT GACCTGGAAT
1001 CCAAGCGGAT ATTAACTTT TTGATATATT TTCACAGAG GTGGCTACAG
1051 TGATACAGTC TAGACAAGAC ATGCTTTCAG AGGATGTTGT ATCTTTACAA
1101 GTCTCTCTGA TTATCTTGG CATGAATGT TACCTGATC GTGTGACTA
1151 TGTGTATAAA GTTCTAGAAA CAACAGTGA GATATTCAAT AAGCTCAACC
1201 TTGAACATAT TGCTACCACT AGTGCAGTTT CAAGGAACCT CACCAGACTT
1251 TTGAAGATAC CAGTTGACAC TTACACAAT ATTTTAACAG TCTTGAATTT
1301 AAAACATTTT CAGCCACTCT TTGATGACTT TGACTGAGG TCCGAAAGA
1351 GCATGAGTTG TTATGTGCTT AGTAATGTTT TGGATTATA CACGAAATTT
1401 GTCTCTCAAG ACCAGGTGGA TTCCATAATG AATTGGTAT CCACGTGAT
1451 TCAAGATCAG CCAGATCAAC CTGTAGAAGA CCTGTACCA GAAGATTTTG
1501 CTGATGAGCA GAGCTTTTGG GGCCTCTTCA TTATCTGCT GGCCTCTGAG
1551 GACCCTGACC AGCAGTACTT GATTTTGAAC ACAGCAGAA ACATTTTGG
1601 AGCTGTGGA AATCAGCGGA TTGCTTTCAC ACTGCCACCT TTGGTATTGG
1651 CAGCTTACCA GCTGGCTTTT CGATATPAAG AGAATTTCA AGTGGATGC
1701 AAATGGCAA AGAATGGCA GAGATTTTTC TCAATTTGCC ACCAGACTAT
1751 CAGTCTTTG ATCAAGCAG AGCTGGCAGA ATTGCCCTTA AGACTTTTTC
1801 TTCAGGAGC ACTAGCTGCT GGGGAATTG GTTTTGAATA TCATGAGACA
1851 GTCCGATATG AATCATCTC CCAGGCATT TTCTGTATG AGATGAAT
1901 CAGCGATTCC AAAGCAGAG TAGCTGCCAT CACCTTGATC ATTTGGACCT
1951 TTGAAGGAT GAATGCTTC AGTGAAGAGA ATCATGAACC TCTGAGGACT
2001 CAGTGTGCCC TTGCTGCATC CAAACTTCTA AAGAACTGT ATCAGGGCCG
2051 AGCTGTGAGC ACCGTGCGC ATCTCTCTG GTCTGGGAGA AACACGGACA
2101 AAATGGGGA GGAAGCTTAC GGAGGCAAGA GGTAAATGGA GTGCTTAAAA
2151 AAGCTCTAA AAATAGCAA TCAGTGCATG GACCCCTCTC TACAAGTCCA
2201 GCTTTTATA GAAATCTGTA ACAGATATAT CTATTTTAT GAAAGGAAA
2251 ATGATGCGT ACAAATCTG GTTTTAAAC AGCTATCTA AAGATTGGA
2301 GAAGACCTCC CGAATCTTGA ATCCAGTGAA GAAACAGAGC AGATTACAA
2351 ACATTTTCA AACACACTGG AGCAATTTGG CTTCGGCGCG GAATCACCAG
2401 AATCCGAGGG GCCAATTTAT GAAGGTCTCA TCCTTAAA AGGAAATAGC
2451 TCACATACT CTTTTCATC TACATCAAT GAGGTTTIA TTACGTAGG
2501 TTTCCCTTCC ATAGATTGTG CCTTTCAGAA ATGCTGAGGT AGGTTTCCCA
```

WO 01/12659

PCT/IB00/01496

2551 TTCTTACCT GTGATGTGTT TTACCCAGCA CCTCCGGACA CTCACCTTCA  
2601 GGACCTTAAT AAAATTATTC ACTTGGTAAG TGTTCAAGTC TTCTGATCA  
2651 CCCCAAGTAG CATGACTGAT CTGCAATTTA AAATTCCTGT GATCTGTAAA  
2701 AAAAAAA

#### BLAST Results

Entry AC007223 from database ENSLHEM:  
Homo sapiens chromosome 16 clone 480G7, WORKING DRAFT SEQUENCE, 38  
unordered pieces.  
Score = 1081, P = 2.8e-217, identities = 219/221  
13 exons

Entry HS015146 from database EMBL:  
human STS W1-8848.  
Score = 2033, P = 2.9e-87, identities = 425/436

#### Medline entries

96327632:  
Genetic mapping and embryonic expression of a novel, maternally  
transcribed gene Men3.  
97258867:  
Endosome to Golgi retrieval of the vacuolar protein sorting receptor,  
Vps10p, requires the function of the  
VPS29, VPS30, and VPS35 gene products.  
92360909:  
Alternative pathways for the sorting of soluble vacuolar proteins in  
yeast: a vps35 null mutant missorts and  
secretes only a subset of vacuolar hydrolases.  
10198044:  
Distinct Domains within Vps35p Mediate the Retrieval of Two Different  
Cargo Proteins from the Yeast  
Prevacuolar/Endosomal Compartment

#### Peptide information for frame 3

ORF from 48 bp to 2435 bp; peptide length: 796  
Category: strong similarity to known protein  
Classification: unset

1 NPTTQSPQD EQEKLIDEAI QAVKVSQFM KRCLDENKLM DSIKHASHML  
51 GELRTSMISP KSYELYNAT SOELHYLEVY LTDEFAGGRK VADLYELVOY  
101 AGNIIPRLYL LITGVVYVYK SFQSRKDIK KDLVEMCRGV QHPLRGLFLR  
151 NYLLOQTRMI LPDLCEPTDE ETTGDISDM DFVLLMFAEM KLMVVRMHO  
201 GSRDREKRE RERQELRIIV GTNLVLSOL EGVNRYVYQ IYVLTGILEQV  
251 VNCRDALAQE YLNECIIOVF FDEFHLOTLM PFLRACAEHL QMNVKNIII  
301 ALIDRIALFA HREDQPGIPA DIKLFDFSQ QVATVIOQRQ DMPSQDVVSL  
351 QVSLINLAKK CYPDRVDVVD KVLCTTVEIF KKLWLEHAT SSVSKELTK  
401 LKEIPVUTYN NLTVLKLM FNPLFEYFDY EDRKSHSCV LSHVLDYNTG  
451 IVSQDQVDSI MNLVSTLIQD QPDQVEDPD PEDFADEQSL VGRFIRLLRS  
501 EDPDQVYLIL NTARKHFGAG GNRIRNFTLP PLVFAAYQLA FRYKENSQVD  
551 DMWKKQKQI FSPKQITSA LKALAEALP LRLFLQALA AGEIGEMNE  
601 TVAYEFMSQA FSLYEDEISD SKAQLAAITL IIGTFERMKC FSEENREPLR  
651 TQCALAASKL LKKPDQGRAV STCAHLFWSG RMTDKNGEEL HGGKRVMECL  
701 KKALKIANOC MDPSIQVOLF IEILNRYIYF YERENDAVTI QVNLQIQKI  
751 REDLPWLESS EETEOINKHF HNTLEHLRLR RESPESEGI YEGULIL

#### BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphtes3\_72p16, frame 3

TREMBL:AF024504.3 gene: "A\_TH017A05.7"; Arabidopsis thaliana BAC  
TH017A05., N = 2, Score = 327, P = 1.9e-162



PIR:S56936 vacuolar protein-sorting protein VPS35 - yeast  
(Saccharomyces cerevisiae), N = 3, Score = 826, P = 1.5e-316

TREMBL:MH47024\_1 gene: "Mem3"; product: "MEM3"; Mus musculus  
maternal-embryonic 3 (Mem3) mRNA, complete cds., N = 1, Score = 3376, P  
= 0

TREMBL:S42186\_1 gene: "VPS35"; product: "Vps35p"; VPS35=vacuolar  
protein sorting [Saccharomyces cerevisiae=yeast, Genomic, 3790 nt], N =  
3, Score = 813, P = 4.4e-115

>TREMBL:MH47024\_1 gene: "Mem3"; product: "MEM3"; Mus musculus  
maternal-embryonic 3 (Mem3) mRNA, complete cds.  
Length = 754

## HSPs:

Score = 3376 (506.5 bits), Expect = 0.0e+00, P = 0.0e+00  
Identities = 666/721 (92%), Positives = 682/721 (94%)

Query: 78 EVYLTFEFAKGRVADLYELVQVAGNIIPRLYLITVGVVVYKSPQSRKDIKDLVEMC 137  
+VYLTFEFAKGRVADLYELVQVAGNIIPRLYLITVGVVVYKSPQSRKDIKDLVEMC  
Sbjct: 34 KVLTFEFAKGERLADLYELVQVSGHIIIPRLYLITVGVVVYKSPQSRKDIKDLVEMC 93

Query: 138 RGVOHPLRGLEFLRNLYLQCTRNILPDEGEPTDEETTGDISDNDVLLNFADHKLWVRM 197  
RGVOHPLRGLEFLRNLYLQCTRNILPDEGEPTDEETTGDISDNDVLLNFADHKLWVRM  
Sbjct: 94 RGVOHPLRGLEFLRNLYLQCTRNILPDEGEPTDEETTGDISDNDVLLNFADHKLWVRM 153

Query: 198 OHQCHSRDRKRRERERQELRLVGTNLVRLSLEG-VHVERYKQIVLTGILEQVVCNCRDA 256  
OHQCHSRDRKRRERERQELRLVGTNLV L+ + -QIVLTGILEQVVCNCRDA  
Sbjct: 154 OHQCHSRDRKRRERERQELRLVGTNLVLTLSVWRCKCQTLQOIVLTGILEQVVCNCRDA 213

Query: 257 LAQEVLMECIIQVFPDFHQLQTLNPFRLACAEHQNVNVMIIIALDRLALFAHREDGP 316  
LAQE KECIIQVFPDFHQLQTLNPFRLACAEHQNVNVMIIIALDRLALFAHRE P  
Sbjct: 214 LAQEVLMECIIQVFPDFHQLQTLNPFRLACAEHQNVNVMIIIALDRLALFAHREDGP 273

Query: 317 GIPADIKLFDIFSOQVATVIQSRQDMPSQVSLQVSLINAMKCYPDVVDVVKVLETT 376  
GIPA++KLFDIFSOQVATVIQSR+DMPSEDVVSQVSLINAMKCYPDVVDVVKVLETT  
Sbjct: 274 GIPAEIKLFDIFSOQVATVIQSRQDMPSQVSLQVSLINAMKCYPDVVDVVKVLETT 333

Query: 377 VEIENKLNLEHIATSSAVSKELTRLLKIPVDTYNNILTVLKLKHPFLFEYFDYESR--K 434  
VEIENKLNLEHIATSSAVSKELTRLLKIPVDTYNNILTVLKLKHPFLFEYFDYES K  
Sbjct: 334 VEIENKLNLEHIATSSAVSKELTRLLKIPVDTYNNILTVLKLKHPFLFEYFDYESRPGK 393

Query: 435 SMSCYVLSNVLNTEIVSQDQVDSIMHLVSTLIQDQPDQVDPEDPQVDEQVSLVGRF 494  
SMSCYVLSNVLNTEIVSQDQVDSIMHLVSTLIQDQPDQVDPEDPQVDEQVSLVGRF  
Sbjct: 394 SMSCYVLSNVLNTEIVSQDQVDSIMHLVSTLIQDQPDQVDPEDPQVDEQVSLVGRF 453

Query: 495 IHLRSDDPQQYLIINTARKHFGAGGNQRIPTLPPLVFAAYQLAFRYKENSXVDXWE 554  
IHLRS+DPQQYLIINTARKHFGAGGNQRIPTLPPLVFAAYQLAFRYKENSX +  
Sbjct: 454 IHLRSDDPQQYLIINTARKHFGAGGNQRIPTLPPLVFAAYQLAFRYKENSXVHTSGK 513

Query: 555 KCKQIFSFHQITISALIKAEALPLRLFLQGLAAGEIGFENHETVAYEFMSQAFSLY 614  
+ \*\* F HQITISALIKAEALPLRLFLQGLAAGEIGFENHETVAYEFMSQAFSLY  
Sbjct: 514 KMRARYFHLHQITISALIKAEALPLRLFLQGLAAGEIGFENHETVAYEFMSQAFSLY 573

Query: 615 EDEISDSKAQAAITLIIGTFERNKCTSEENHEPLRTQCALAASKLLKKPQGRAVSTCA 674  
EDEISDSKAQAAITLIIGTFERNKCTSEENHEPLRT+CALAASKLLKKPQ C  
Sbjct: 574 EDEISDSKAQAAITLIIGTFERNKCTSEENHEPLRTQCALAASKLLKKPQQAERHMTCT 633

Query: 675 HLPWSGRNTDKNGEELHGGKRVNCLKALKIANQCMDFSLQVLFIEILNRYIFYEKE 734  
L WSGRNTDKNGEELHGGKRVNCLKALKIANQCMDFSLQVLFIEILNRYIFYEKE  
Sbjct: 634 SL WSGRNTDKNGEELHGGKRVNCLKALKIANQCMDFSLQVLFIEILNRYIFYEKE 692

Query: 735 NDAVTIOVLNQLQIKTREDLPNLESSEETEQINKHFNTLEHLRRSPESSEGPIVEGL 794  
NDAVTIOVLNQLQIKTREDLPNLESSEETEQINKHFNTLEHLRRSPESSEGPIVEGL  
Sbjct: 693 NDAVTIOVLNQLQIKTREDLPNLESSEETEQINKHFNTLEHLRRSPESSEGPIVEGL 752

Query: 795 IL 796  
IL  
Sbjct: 753 IL 754

Pedant information for DKFZphtes3\_72p16, frame 3

Report for DKFZphtes3\_72p16.3

(LENGTH) 796

930

**PCT/IB00/01496**

(No Prosite data available for DKFZphtes3\_72p16.3)  
(No Pfam data available for DKFZphtes3\_72p16.3)

DNFzphes3\_7b22  
-----

group: cell structure and motility

DNFzphes3\_7b22 encodes a novel 443 amino acid protein with weak similarity to paramyosins.

The novel protein is related to paramyosin, a major structural component of thick filaments and invertebrate muscle. Paramyosins are promising antigens for immunization against several parasites, such as *Schistosoma mansoni*.

The new protein can find application in modulating cell adhesion/motility and membrane/cyto skeleton structure and dynamic.

similarity to paramyosins

complete cDNA, complete cds, few EST hits

Sequenced by BNF2

Locus: /map="3"

Insert length: 2291 bp  
Poly A stretch at pos. 2241, polyadenylation signal at pos. 2213

```
1 GGAAGAAAGG CTAGCGGGCG TTGCGCGTAT GTGGGTGCTT TGAGGCAGTT
51 TTTCAGTTCT TTCAATTACC AAAGTGACAT GCACCTACTA GGTGCGAGGT
101 GTTTGACGCT ACATACAACC CTCTGCAAAA TCTTTTCAGT TAGTCTCTCG
151 TATGAAAAGT TTCCAGCCAA GAATGGCCAC TGCACCTGAG ATAAGGGGGA
201 TCCTGGCCAT TAAGGAACCC TTGCTTTCGA AACTGAGCCG TGAGGAACCTA
251 TACAAATGGG GAATTTGGGA CAATCCGAC TGCTCATGTA CACTAAGAAG
301 TAAATATTACG AACTCACTGA GCTGGAAGTC ATTCAACGGG AATTGAATAG
351 GTAAGTGCAC TTTTGTGAGA TTATAAATAT ACCACGGAGG GTAACGAAGC
401 TACAGAAGAA TGGAAAGAAG CAGCTCTGAA GACTCAAAAC TTCTCCAAA
451 AGTTTGCAAT TCTGAGATGA CGGTGTCAGT GACAGCGCAA CCACCTAGTA
501 CCGTAGAAGA AGAAGGAATA CCTAAAGAAA CAGACATAGA AATCATCCCA
551 GAAATCCCGG AAACCTCTAGA GCACTGTGCC CTTCAGATG TGCTGAGGAT
601 CTGCGCAGCT CTGAGGACA CACAGACACA GCTCTCTATT TGAACATGA
651 TCATGCCCGT TCAGTACGAA GGGAGACAGA GCATCTGCGT GAAAAGCAGA
701 GAAATGAATC TAGAAGGAAC GAATCTAGAC AAATTCCTCA TGGCTCAAC
751 AATCACAATA ATACCCAGTC CGTTAATAGC TGAGGAAGGA CCCAATTGCG
801 CAGAAATCAG ACACAGAGGC CGGTTCGCTG TGCAGTTTAA CAATAATGAG
851 GATCTGTCTC TCAGAAACCC TACAGGCGAC ACCATCATGA CTACGGAGAC
901 ACTGAAGAAA ATTGAGATTG ATAGGCAATT TTTACGCGAT GTGATTGCG
951 ATACCATYAA GGAATTGCAA GATTGCGCCA CTTCACAGAG TCCTCTGCAA
1001 GCTTTGAGCA AAGGAGGGA AACCAAAATG CATTTCTATG ACATCAATTG
1051 CAGGAGGAAA AAAGGAAGAA AACAGATAAT ATCACTTCAA AAACAGCTAA
1101 TTAATGTCAA AAAGGAATGG CAATTTGAAG TCCAGAGTCA GAATGAGTAT
1151 ATGCTTAACC TCAGAGGACA ACTGCAAGAG ATGAAGGCAA ATCCCACTT
1201 GGAGATCGC TACATGAAA CCAATACCGA GCTGCAGATT GCCCAGACCC
1251 AGAATAAGTG TAACAGAACA GAGGAACCTC TGGTGGGAAG GATTGAGAAA
1301 CTCAGATGTA AAACCGAAGA AGAGGCCCGG ACTCATACAG AGATTGAAAT
1351 GTTCTCTAGA AAGGAGCAG AGAAGCTTGA GCAAGGCTG GACTCTGCA
1401 TGGAGAAATA CGATAAGGAC ACAGAAATGA AACAGAAATG ACTAATGCT
1451 CTCAGAGCCA CAAGGCCAGC TGACTTAGCA CACCTTCAAG ACCTGGCAAA
1501 GATGATAGTA GATATGAAAC AGGTATCATC TGAAGATGCT ATAGAAAGGG
1551 AGAGGAGCAA GAGAGAGGTA AAGCAGGATC TCTTGGAAAT AAAGAGCCTT
1601 ATAAAGCTCC AGGCCTGGTG GCGAGGCACT ATGATACGGA GAGAAATTGG
1651 TGTTTTCAGG ATGCTTAAGG ACAAGATTGA TAGCAGGAT TCAGAAAGGCA
1701 AAGTTAAGG CAGGATAGG AGGAGAGGGA AGAAGAGTGT ACCAATGCT
1751 CTTTTGTGTT TTCTGCTGTT ATTCTGGAGG TGGGAAGGAC TTGAGAGATT
1801 AAGAAGACCC TGGTACCTCA AAGATGACTC ATCTACAGGT TGTTCCTAT
1851 TGAGACTTTC CCAGGGAAGC CTGATTTCAC TTGCTGTTT AATTTCACCT
1901 TGCTGTGTTG GTGGTTTTTC AAACCTGATC TTAGATATAC ACATATGACT
1951 TAGGGCTTCC TCATACCTTG CTGGGAAGAA GTTCTAGTA GTCTGTGAA
2001 GATTCTTCT TCTGCTCTTT TCTCAGAGA ACAAGGAGT TCACCTGGCTT
2051 AGCTACAGTC ACCCATTTGA ACTTGAGTAA TTCTGTATAT GTCAGATTTT
2101 GATTTTACCT AATTGCTCTC TACTGAAAAA ACTCTTAGTA GCAGAAATAT
2151 TCAGTAGGAA TTACAATATG ATGTTATTAG CTGTCAGCA TAATATATAC
2201 ACAGCAAGAT TTTAATAAAT GTTGGTCTCT GCTGCTCTTT TAAAAAATAA
2251 AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA A
```

BLAST Results  
-----

Entry G36731 from database ENBL:  
SHGC-52923 Human Homo sapiens STS cDNA.

WO 01/12659

PCT/IB00/01496

Score = 2262, P = 1.3e-97, identities = 462/468

Medline entries

No Medline entry

Peptide information for frame 2

ORF from 410 bp to 1738 bp: peptide length: 443  
Category: similarity to known protein

```
1 MEEDSLEDN LPFKVHSEM TVSVTGEPPS TVEEGIPKE TOIGIIPFIP
51 ETEPLESLPD VLRIASVLED TTDQLSILMY IMPVOYEGRO SICVKSREDM
101 LEGTNLDKLP MASTITKIP9 FLITEEGPWL PEIRHGRITA VFNKMDOLV
151 FKFPFQZIM TYTELKIQI DROFFSOVIA DTIKELQDRA TVNSLLQALS
201 KERENKMHFY DIAREEKGR KQIISLQKOL INVKKEMQTE VQSQNEYIAN
251 LKDLQENKA KSNLENAYNK THTELQIAQT QKCKRTEEL LVEETELRM
301 KTEEDARTHT ELENFLAREQ QLEEDKLEPW MERYDKDEM KQNELNALRA
351 TIASDLAWLQ DUAHMIREY QVITEERIEK ESKKVKQO LLEKSVIKL
401 QAWWRGTMR REIGGFMPK DKVDSKDSKG KGGKDKRRG KKK
```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DRFzptes3\_7b22, frame 2

SWISSPROT:MYSP\_BRUMA PARAMYOSIN., N = 1, Score = 158, P = 5.8e-08  
PIR:A44972 paramyosin - nematode (Dirofilaria immitis) (fragment), N =  
1, Score = 157, P = 7.1e-08  
SWISSPROT:MYSP\_OMCVO PARAMYOSIN., N = 1, Score = 157, P = 7.4e-08  
PIR:S52537 emm L 15 protein - Streptococcus pyogenes, N = 1, Score =  
151, P = 8.6e-08

>SWISSPROT:MYSP\_BRUMA PARAMYOSIN.  
Length = 880

HSPs:

Score = 158 (23.7 bits), Expect = 5.8e-08, P = 5.8e-08  
Identities = 66/259 (25%), Positives = 125/259 (48%)

Query: 142 EFNKQQLVFKKPTQIMTTETLKKIQRQFFSDVIADTIKELQDSATYNSILQALS 201  
+ K + L R R T E K++ + +D +A + LQ A N LL+ +  
Sbjct: 169 QLKXDNLAEKAAERFEATVLSNKKVEDLNHVNQ-LAQQORLQ--AENDLLKEIND 225  
Query: 202 ER---ENNMHF-YDIIAREEKGRKQIISLQKOLINVKKEMQFEVOSONEYIANLKQQLQE 257  
++ +N H Y + +E+ R+++ +++ ++ + +VQ + + + D+ E  
Sbjct: 226 QKVQLDNLQHVYOLAQOLEEARRRLEDAERERSQLQAQLH-QVQLELDSVPTALDE--E 282  
Query: 258 MKAKSNLENRYNKTHTELQIAQTQKCKRTEELVVEEIEKLMKT-EZEARTHTIEHFL 316  
A++ E++ NTE I Q + K + L EE+E LR K +++A +IE+ L  
Sbjct: 283 SAARAEEHKLALANTE--ITQWKSFKDAEVALHHEEVEDLRKMKLQKQAEYEEQIEIML 340  
Query: 317 RKEQO--KLEERLEFVNEKYDKDTEMKQNELNALKATKASDLAHLQDLAKMIREYEQVII 374  
+K Q K + RL+ +E D E QN + L+ K + L K + E + I  
Sbjct: 341 QKISOLEKAKSRLOSEVEVLVDLEKAQNTIILERAK-----EQLEKTVNELKVRID 393  
Query: 375 EDRIKERSKKRVKQOOLEKSVIKL 400  
E +E E +++ + L EL+ + L  
Sbjct: 394 ELTVELEAAQREARAAALELQKLKHL 419

Score = 118 (17.7 bits), Expect = 1.3e-03, P = 1.3e-03  
Identities = 54/231 (23%), Positives = 108/231 (46%)

Query: 181 DTIKELQDSATYNSLIQ----ALSKERENKMHFYDIIAREEKGR-RKQIISLQKOLINVKK 235  
D +KE+ D LQ L+++ E + RE + Q+ +Q +L +V+  
Sbjct: 218 DLLKEINDQKVLQDNLQHVYOLAQOLEEARRRLEDAERERSQLQAQLHVOLELDSVRT 277

Query: 236 EVOFE--VQSNEY-IAHLKDLQGHKAKSLNRYNKTNT-LOIAOTOKKCHTEELL 291  
E + + + E + + A + + + K + K + E E L + OK + E + +  
Sbjct: 278 ALDESAARAAEAHLALANTEITOMKSKFQAEVALHHEEVEDLRKMLQKQAEYEEQIE 337  
Query: 292 VEEIEKLKMTTEEARNTTEIMF---LRKEQKLE--ERLEFNMKEYDKTENKQNELM 346  
+ + + K + + + R + E + E + L K Q + ER + + EK + + + + EL  
Sbjct: 338 IM-LQKISQLEKAKSLQSEVEVLVDLEKQNTIATLERAKQLEKTVNLEKVRIDELT 396  
Query: 347 A-LKATKASDLAHLQDLAKMIREYQVIEEDRIEKSKKKVKQDLLEKSVI 398  
L + A + A L + L K + YE + E + R KK + + DL E K +  
Sbjct: 397 VEEAAGREARAAELQKLNLYEKAV-EQKEALARENKQLQDLHEAKAL 448  
Score = 107 (16.1 bits), Expect = 2.1e-02, P = 2.1e-02  
Identities = 49/279 (17%), Positives = 124/279 (44%)  
Query: 123 IYEGPHLPIRHHGRFAV-EFNHMOVLKPKPTROTINTTFLKKIQIDRQFFSDVIAD 181  
I E L + R A + E K + + L E + + E KK + Q D + + AD  
Sbjct: 392 IDLTVEEAAQREARAAELQKLNLYEKAVEQKEALAREN-KKQDLHEAKALAD 450  
Query: 182 TIKELQDSATYSLGLQALSKEKHNKHFDYIAREKGRKQ--IISLQKOLNVKKEWOF 239  
+ + L + N + L + E + + + R + R Q + LQ + I + + + Q  
Sbjct: 451 ANKKHLEDLENARLAGEIRELQALKESAAARRDAENRAQALAEQQLRIEHERRLQE 510  
Query: 240 EVQSNEYIANLKDQLEKAKSLNRYNKTNTLOIAOTOKKCHTE-ELLVEIEKL 298  
+ + N + + + A L + + E + + + E E + V + + +  
Sbjct: 511 KEENEALRKNMFEIDRLTAA--LADAKARKAEISRLKRYQAKIAELDHTVDHLRA 568  
Query: 299 RMTTEEARNTTEIMFLEKQKLEERLEFNMKEYDKTENKQNELMALKATKASDLAH 358  
+ + + + + E L + + + + L + + + + + Q + + + AL A + + +  
Sbjct: 569 NIEAQYIKKQSEQLK:LOASLEDTQKQQLDQY----ALAQKVSALSA-ELEECKV 623  
Query: 359 LQDLAKMIREYQVIEEDRIEKSKKKVKQDLLEKSVIKIQ 401  
D A R + + + E + + V + L + K + + + +  
Sbjct: 624 ALONAIRAKQAEIDLEEANGRITOLVSVNHNITATNKKLETE 666

Patent information for DKF2phtes3\_7b22, frame 2

Report for DKF2phtes3\_7b22.2

[LENGTH] 443  
[MM] 51917.95  
[PI] 6.18  
[HOMOL] PIR:S28589 trichohyalin - rabbit 2e-08  
[FUNCAT] 30.03 organization of cytoplasm [S. cerevisiae, YDL058w] 7e-07  
[FUNCAT] 08.07 vesicular transport (golgi network, etc.) [S. cerevisiae, YDL058w] 7e-07  
[FUNCAT] 1 genome replication, transcription, recombination and repair [M. jennaschii, MJ1322] 5e-06  
[FUNCAT] 03.22 cell cycle control and mitosis [S. cerevisiae, YPR141c] 1e-05  
[FUNCAT] 03.13 meiosis [S. cerevisiae, YPR141c] 1e-05  
[FUNCAT] 11.01 stress response [S. cerevisiae, YPR141c] 1e-05  
[FUNCAT] 03.07 pheromone response, mating-type determination, sex-specific proteins [S. cerevisiae, YPR141c] 1e-05  
[FUNCAT] 08.22 cytoskeleton-dependent transport [S. cerevisiae, YPR141c] 1e-05  
[FUNCAT] 09.10 nuclear biogenesis [S. cerevisiae, YPR141c] 1e-05  
[FUNCAT] 30.05 organization of centrosome [S. cerevisiae, YPR141c] 1e-05  
[FUNCAT] 06.10 assembly of protein complexes [S. cerevisiae, YPR141c] 1e-05  
[FUNCAT] 99 unclassified proteins [S. cerevisiae, YOR216c] 3e-05  
[FUNCAT] 11.04 dna repair (direct repair, base excision repair and nucleotide excision repair) [S. cerevisiae, YER095w] 6e-05  
[FUNCAT] 30.10 nuclear organization [S. cerevisiae, YER095w] 6e-05  
[FUNCAT] 30.02 organization of plasma membrane [S. cerevisiae, YER008c] 1e-04  
[FUNCAT] 08.16 extracellular transport [S. cerevisiae, YER008c] 1e-04  
[FUNCAT] 03.04 budding, cell polarity and filament formation [S. cerevisiae, YER008c] 1e-04  
[FUNCAT] 30.04 organization of cytoskeleton [S. cerevisiae, YDR356w] 2e-04  
[FUNCAT] 08.01 nuclear transport [S. cerevisiae, YDL207w] 4e-04  
[FUNCAT] 04.07 rna transport [S. cerevisiae, YDL207w] 4e-04  
[FUNCAT] 06.07 protein modification (glycosylation, acylation, myristylation, palmitoylation, farnesylation and processing) [S. cerevisiae, YKL201c] 5e-04  
[EC] 3.6.1.32 Myosin ATPase 3e-08  
[PIRKM] phosphotransferase 6e-06  
[PIRKM] citrulline 8e-06  
[PIRKM] tandem repeat 1e-07  
[PIRKM] heart 6e-06  
[PIRKM] polymorphism 4e-06  
[PIRKM] serine/threonine-specific protein kinase 6e-06  
[PIRKM] DNA binding 8e-08

[illegible]

PS00001	285->289	ASN_GLYCOSYLATION	PDCC00001
PS00004	152->156	CAMP_PHOSPHO_SITE	PDCC00004
PS00005	164->167	PKC_PHOSPHO_SITE	PDCC00005
PS00005	182->185	PKC_PHOSPHO_SITE	PDCC00005
PS00005	280->283	PKC_PHOSPHO_SITE	PDCC00005
PS00005	383->386	PKC_PHOSPHO_SITE	PDCC00005
PS00006	5->9	CK2_PHOSPHO_SITE	PDCC00006
PS00006	30->34	CK2_PHOSPHO_SITE	PDCC00006

WO 01/12659

PCT/IB00/01496

PS00006	41->45	CK2_PHOSPHO_SITE	PDOC00006
PS00006	57->61	CK2_PHOSPHO_SITE	PDOC00006
PS00006	104->108	CK2_PHOSPHO_SITE	PDOC00006
PS00006	182->186	CK2_PHOSPHO_SITE	PDOC00006
PS00006	243->247	CK2_PHOSPHO_SITE	PDOC00006
PS00006	262->266	CK2_PHOSPHO_SITE	PDOC00006
PS00006	271->275	CK2_PHOSPHO_SITE	PDOC00006
PS00006	302->306	CK2_PHOSPHO_SITE	PDOC00006
PS00006	308->312	CK2_PHOSPHO_SITE	PDOC00006
PS00006	310->314	CK2_PHOSPHO_SITE	PDOC00006
PS00007	261->269	TYR_PHOSPHO_SITE	PDOC00007
PS00007	184->193	TYR_PHOSPHO_SITE	PDOC00007
PS00009	218->222	AMIDATION	PDOC00009
PS00009	439->443	AMIDATION	PDOC00009

(No Pfam data available for DKFZphtes3\_7b22.2)



DKFZphtes3\_7d17

group: testis derived

DKFZphtes3\_7d17 encodes a novel 633 amino acid protein with weak similarity to human KIAA0454.

Pfam predicts a TNFR/MGFR cysteine-rich region.

No informative BLAST results: No predictive prosite or SCOP motifs.

The new protein can find application in studying the expression profile of testis-specific genes.

similarity to KIAA0454

complete cDNA, complete cds, EST hits

Sequenced by BMFZ

Locus: unknown

Insert length: 3608 bp

Poly A stretch at pos. 3587, polyadenylation signal at pos. 3570

```
1 GGGAAAGTTAC GGGCAAGTCC ACCCAGCGTT TCTCAGGCAA TCTGAAGGCA
51 AATCCTGTTT AGACCCAGGC GAAGGTTCTT GGTGACCCAG GCTCTCACCA
101 GCCAATTGTC CTTGCCGCTC CTCTGAGGGT TATCTGGAGC TTCAGTGGTG
151 TGTGCTCTTG GCTCCACACG TGGGGAATGG ACTGATGCC ACCTGCGAGG
201 GCTTCAGTGT GACTCTCCGA GGCCCTGATG TAGAACTTC CCCATTCGGT
251 GCACCAAGAG CAGCCTCACA TGGTGTGGGG CGACATCAAG AGCTGCGAGA
301 TGCAACAGTC CTTGCCGCCA CCTCTTCTGG CACAAGAGTC AGCATGGTGG
351 TATCTGGCGG CCTCTGGTCC GTTGAGAGAG CAGAGATGAA CATCTAGCAA
401 ATCAACAGAA AATCGGCGCC CCAGCTGGCA GAGAACAAAC AGCAGTTCAG
451 AAACCTCAAA CAGAAATGTC TTGTAACCTA AGTGCCCTAC TTCTGGGCCA
501 ACCCGCAAAA TATTTAGAGC TATGAGAGCT GCAGAGAGCT CATTAATCTT
551 ATGCTGAGGG ATGAGCGGCT GCTCAGAGAA GAGAAGCTTG CAGAGGAGCT
601 CGGGCAAGCT GAGGAGCTCA GGCAATATAA AGTCTGTGTT CACTCTCAGG
651 AACGAGAGCT GACCCAGTTA AGGAGAGAGT TACAGGAGGG CAGAGATGCC
701 TCCCGCTCAT TGAATCAGCA TTCTCAGGCG CTCTCTACTC CGATGAGGCC
751 GGACAACTCC CAGGGAACGG ACCTCCGAGA ACAGCTGGCT GAGGGATGTA
801 GGCTGGCACA GCACCTGCTC CAAAAGCTCA GCCCAGAAAA TGATGACGAT
851 GAGGATGAAG ATGTTAAAGT TGAGGAGGCT GAGAAAGTAC AGGAATTATA
901 TGCCCGCAGG GAGTGCCAGA AGGCTGAAGA AAAGGAAGTC CCTGAGGACT
951 CACTGGAGGA GTTGCCCATC ACTTGTTCAA ATAGCCACCA CCCTGTGTAG
1001 TCCAAACGAG CTTACGGGAA CACCAAGATC ACATTTGAGG AAGACCAAGT
1051 CGACTCAACT CCAATTGACT CATCTCTCTA TGATGAATGG TTGCAATGCT
1101 TATGCATTAT CCCAGAAAT GAAAGTGATC ATGAGCAAGA GGAGAAAAAA
1151 GGGCCAGTGT TCCCAAGGAA TCTCAGGAG TCTGAAGAGG AGGAAGCCCC
1201 CCAGGAGTCC TGGGATGAAG GTGATTGGAC TCTCTCAATT CCTCTGACA
1251 TGTCTGCTCT ATACCAATCT GACAGAGACA CCTTCACTC AGTAGAGGAA
1301 CAGCAAGTCC GCTTGGCTCT TGACATAGGC AGACATTGGT GTGATCAAGT
1351 GAAAAAGGAG GACCAAGAGG CCACAGTCC CAGGCTCAGC AGGAGCTGCT
1401 TGGATCAGAA AGAGCCTGAA GTTCTCAGG ACTCACTGCA TAGATTTTAT
1451 TCAACTCCTT TTGAGTACCT GGAAGTCCCT GACTTATGCC AGCCCTACAG
1501 AAGTGACTTT TACTCATTCG AGGAACACA CTTTGGCTTG GCTCTTGACT
1551 TGGACGAAAT GAAAAAGGAC CAGAGAGAGG AGAGAGCCA AGGCCACCA
1601 TGCCCGAGGC TCAGCAGAGA GCTGCCGGAG GTAGTAGAGC CTGAGGACTT
1651 GCAGGACTCA CTGGATAGAT GGTATTCCAG TCCTTTCACT TATCCAGAAC
1701 TGCCGTGATC ATGCCAGGCC TACGGAGATT GCTTTTACTC ATYGAGGAA
1751 GAACAGTTG GCTTTTCTCT TGAGCTGAT GAAATTGAAA AGTACCAAGA
1801 AGGGGAAGAA GATCAAAAGC CACCATGCC CAGGCTCAAC CAGGTGCTGA
1851 TGGAGCAGA AGAGCCTGAA GTCTTGAGG ACTCACTGGA TAGATGTTAT
1901 TGCACTACTT CAAGTTACTT TCAACTCAT GCCTCATGCC AGCAGTACAG
1951 AAGTGCTTTT TACTCATTTG AGGAACAGGA CGTCAGCTTG GCCCTTGACG
2001 TGGACAATAG GTTTTTTACT TTGACAGTGA TAAGGCACCA CCGGCCCTTC
2051 CAGATGGGAG TCATATTCCC ACACTAAGCA GGCCTTACTA AGCTGAGAGA
2101 TGTCAATGCT CAGGCGAGGA CCTATAGGCA CATGTAGGTT TGAATGAAGC
2151 TGTAGTCCC TTTGGAAGCC CAGCTATAG ATGGGAAAGT GGGCATGGCT
2201 CTATTCTTAT TCTCAGACCA TGCCAGTGGC CACCTGTGCT CAGTCTGAAG
2251 ACCTTGGACC CAGTTAGGTT GTGACAGCTT CACACACTA TGTAGCAGAT
2301 GCGGGAGTGT ATCTGCAGA CATCTAATT TGAACAGAT ATCTTGGGTT
2351 AGCTACAAAG TTCTCAGGG GTTTCATTTT GCAGGCATGT CTCTGAGCTT
2401 CTATACCTGC TCAAGGTCAG TGTCACTTTT GTGTTAGCT CATCCAAAGG
2451 TGTTACCTGT GTTTCATTTA ACCTAACCC ATTCTGTGTA TCTCTAGTGT
2501 TGCTTTGTTT TACTGATCC ATCTGTACA CAGGAGGAT CTTGGCTGTA
2551 GGATTGTATT TCAGAACCCG TGACTGCTCT TGACAGTTGT TAACCCACTA
2601 GGCTCCTTTG AGTAGAGAAG CCAATGCTCT TCAGCTTCCA ATTGATATCA
2651 ATACTTAGGA AGACCACAGE TAGACGACCA ACAGCATTTG GAGGCGCTTA
```

```
2701 GTCCTGCTCC TTTCATTC ATCTGTAA GAACAGGAGT CAGGAGCCGC
2751 TGGCAGAGA CAGCATGTCA CTTGGGACTC TGCAGTCA GAATATGAAC
2801 AATGCCATGT TCTTGAGAA AATGCTTAGC CTGAGTTTCA TAGAGGTAA
2851 TCACCAGACA ACTGCAGAA GTAGAACACT GAGCAGGACA ACTGACCTGT
2901 CTCCTTTACA CAGTCACGT CACCAGAAAT CACACAACAA AAGGAGGAG
2951 AGATATTTTG GTTTCAGAG AGTAAATGA TAATGTAGCT ACATTTCTTT
3001 ACTATTTTGG ACCGCCAAAT ATTTCCTCAT CTTTGTGTG TTGTATTGA
3051 TTTTGTGAC ATGGACTTGT TTOTAGAGA CAGTCAGCT GTCTGGCTCA
3101 ATGGTCTACA TTCTGAAGTT GTCTGAAAT GTTTCATCA TTAATTCAG
3151 CCTAACGCTT TCATCAGAA CACTCAGAG TCGATACTGT GAGTTTCCAA
3201 CCTCAGCCCA TCTGTGGCCA GAGAAGTCT AGTTTGTCCA TCAGATTAT
3251 CATGATATCA GGACTGGTTA CTTGTTAAG GAGGGGTCTA GGAGATCTGT
3301 CCGTTTACA GACACTTAC TTATGATGA GTATTGGGA GAGTGGTTT
3351 TCAAGTAGA AATGCTCTGT ATTCCAGTGA TCATCCTCTA AAGCTTTAT
3401 CATTATTAA TCATCCCTGC CTGTGCTAT TATTATATC ATATCTCTAC
3451 GCTGGAATTT TGTGCTCTCA ATGTTTACTG TGCTTTTGT TTGCTAGTG
3501 TGTGTGTGTT AAAAAAANC ATCTCTGCG TGAGTTTAA TTTTGTCCA
3551 AAGTATTATT AATCTATACA ATTAANAAT TTTGCTATC AAAAAAANA
3601 AAAAAANA
```

## BLAST Results

No BLAST result

## Medline entries

No Medline entry

## Peptide information for frame 2

ORF from 176 bp to 2074 bp; peptide length: 633  
Category: similarity to known protein

```
1 MFLTPVQGF QNTLRGPDVE TSPFCAPRAA SHGVGRHQEL RDPTVPGPTS
51 SATNVMVVS AGPWSGEKAE MHILEINKS RPLAENKQO FRNLKOKCLV
101 TVAVYFLAAR QNVVDYEDCK DLIKSHADE RLITEELAE ELQAEELRQ
151 TVLVRSQER ELTQLEKIQ EORDASRLH QHLQALLTD EPDMSQGRDL
201 REQLAEGCRL AQHLVQKLSR ENDDDEDVDV KVEAEKVOE LYAPREVOKA
251 EEKEVPEDSL EECATYCSNS RHPCESNQPY QNTRITFEED QVSTLIDSS
301 SHQWLDANC ITPENSUNE CEKEKQPSR RLQSEEEZEE APQSDWGDG
351 WTLSPFPMDS ASYQSDRSTF NSVEEOOVLG ALDIGRHMCD QVKKEDQDAT
401 SPRLRELLD EKEPEVLQDS LDRFYSTPFE VLELPOLCOP YRSDPYSLOK
451 QHLGLALDLD RKKKQCEEEE DQCPKCPALS RELPEVPEFE QLOSLDWHY
501 STFPSPFLSL DSCOPYGCTF YSLEEHVGF SLOWDEIKY OEGEEQKPP
551 CPRLNEVLNE ADEPEVLQDS LDRCYSTTST YFOLHASFOQ YRSAPYSFEE
601 QOVSLALDVO NRFTLTIVR HHIAFQMGV1 FPH
```

## BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphtes3\_7d17, frame 2

PIR:T00069 hypothetical protein KIAA0454 - human (fragment), N = 1,  
Score = 199, P = 1e-11

PIR:A45592 liver stage antigen LSA-1 - Plasmodium falciparum, N = 1,  
Score = 158, P = 2.7e-07

>PIR:T00069 hypothetical protein KIAA0454 - human (fragment)  
Length = 1,882

## HSPs:

Score = 199 (29.9 bits), Expect = 1.0e-11, P = 1.0e-11  
Identities = 74/261 (28%), Positives = 122/261 (46%)

Query: 117 EDCKDLIKSHLDERLLT---EEKLAELGQAEELAQYKVLVHSERELTLQREKIQEG 172  
+D + LI+ + + E L EEKLAEL A +Y L+ Q REL+ LR+K+EG  
Sbjct: 964 KDLESLIQRVSQLEAQLPKNGLKEELAEELRSASWPKGYDSLIQQAELSYLRKIRREG 1023

Query: 173 RDASRLNHQ-----LQALLTPDEPDSQGRDLREQLAEGCRLAQHLVOKLSFENDD 225  
R + +H + LL + D G+ REQLA+G +L + L KLS +  
Sbjct: 1024 RGICYLITRAKOTVKSFDLLKSNIDITLQSPFREQLAGSQLTERLTSKLSKDHKS 1083

Query: 226 EDEDVKVEAEKVOLYAPREVQAEER-EVPEOSLEECATCSNHNPCSHQPTGHTA 284  
E + +E L RE+Q+ E+ EV + L+ +T S+SH +S++ +T  
Sbjct: 1084 EKDAQGLEPLA----LRLSRELQEKKEVIEVQAKLDARSLTPSSSHALSDSHRSPTS 1139

Query: 285 ITFEEDQV--DSTLISSSDHDLDAVCIIPENESDHEQEKEKGPVSPNQLQSEEEAP 342  
+E + D + + +H E A P + +S + S + A  
Sbjct: 1140 FLSDLEACSDMDIVSEYTHYEKKAS---PSHSDSIHSHSASVLSKPSSTASQGA 1196

Query: 343 QESWDQEGOMTSLIPPDSASYSQSRSTFH 371  
ES + L P + S FH  
Sbjct: 1197 AES-WSNPISLPTPONTPEANQANSGFH 1224

Score = 89 (13.4 bits), Expect = 1.1e-01, P = 1.0e-01  
Identities = 35/89 (39%), Positives = 44/89 (49%)

Query: 464 KQOEEDQDQ---PCFRLSRELPEVVEP-EDLOSLDRWYSTPSPYLPDSCQ-PYGS 518  
KD + E+DQ P RLSREL E + E LQ LD TP S L DS + P +  
Sbjct: 1079 KDHSEKDAQGLEPLALRLSRELQEKKEVIEVQAKLDARSLTPSSSHALSDSHRSPTS 1138

Query: 519 CFYSLEEHVGFSLDVEIERVQEGEEDQRP 550  
P S E E D+D + + EE + P  
Sbjct: 1139 SFLSDELEACS---DMDIVSEYTHYEKKAS 1167

Score = 73 (11.0 bits), Expect = 4.8e+00, P = 9.9e-01  
Identities = 31/88 (35%), Positives = 40/88 (45%)

Query: 390 DQVKKEDQATSP---RLSRELLD-EKEPEVLQDSLDRFYSTPFEYLELPOLCO-PYRS 444  
D + +DQ P RLSREL + EK EVLQ LD TP L D + P +  
Sbjct: 1080 DHKSEKDAQGLEPLALRLSRELQEKKEVIEVQAKLDARSLTPSSSHALSDSHRSPTS 1139

Query: 445 FYSLQELHGLALDOLRMKKDQEEEDQRP 475  
F S L D+D + + EE + P  
Sbjct: 1140 FLSDLEACSDMDIVSEYTHYEKKAS 1167

Score = 68 (10.2 bits), Expect = 1.1e-01, P = 1.0e-01  
Identities = 36/156 (23%), Positives = 68/156 (43%)

Query: 31 SHGVGRHOELRDPV---PPTSSATNVSMVVSAGPWS-----GEKAEHWILEINK 79  
S G +HQE + TV P P S + V A G++ ++ +  
Sbjct: 684 SPCKHNOHSEGNVTVPFPAPQSLGATTVDAHQDNQSQPRPGQSAFSLPGSTON 743

Query: 80 SRPOLAENKQOFNKLQKCLVTOYAYFL-ANRQNYDYE-DCKDLKSLRDERLLTEER 137  
R OL++ Q++++L++K L+++ F AN Y + L+K + ++ ++  
Sbjct: 744 LRSQSQCKQRYQDLQERLLSEATVFAQAELEKRYHVLGTGESLVKQDSRQIQVQLQDL 803

Query: 138 LAELGQAEELRQY-LVHSQERELTOLREK-LQEG 172  
E G++E + + + E L+E L EG  
Sbjct: 804 GYETCGRSEKAEETTSPECEKNSLKENVLEG 839

Score = 65 (9.8 bits), Expect = 2.2e-01, P = 2.0e-01  
Identities = 23/96 (23%), Positives = 52/96 (54%)

Query: 123 IKSLRDERLLTEELAEELGQAE-----LRQYKVLVHSQERELTOLREKLQEGDASRS 178  
++ + D+ + E + E+ EE LRQ ++ V ++ +L +L+ L ++ +  
Sbjct: 5 LQRINKAKVALERATQENFSALEEKELPOLKLAVERKHOLELRLDVL-----SNEA 60

Query: 179 LNHQLQALLTPDEPDSQGRDLREQLAEGCRLAQHLVOKL 218  
Q ++LL ++G ++ EQL+ C+ Q L +++  
Sbjct: 61 THQSMESLL-----RAKLEV-EQLSTTCQLQLKLEN 93

Score = 61 (9.2 bits), Expect = 3.5e-01, P = 4.2e-01  
Identities = 27/95 (28%), Positives = 47/95 (49%)

Query: 134 TEER-LAEELGQAEELRQY---KVLVHSQERELTOLREKLQEGDASRSLNHQLQALL 188  
+E K L -LG+ EE R Y +LV + + + L+ +Q ++L ++L  
Sbjct: 855 SEARKLENLQKQEEFRVYKSEMILV--LRKDILKQALQANAKVIQHLKSRVSLV 912

Query: 189 PDEPDSQGRDLREQLAEGCRLAQHLVOKLSFENDDDEE 228  
+ +S R R+ A G ++ SP + DEDE  
Sbjct: 913 TSDYSSSLERP-RKLAAGT-----LEGSSPHSVPEDE 945

Score = 57 (8.6 bits), Expect = 1.4e+00, P = 7.5e-01  
Identities = 26/92 (28%), Positives = 47/92 (51%)

Query: 127 LRDRLITEELAEELGQAEEL---RQYKVLVHSQERELTOLREKLQEGDASRSLNHQL 183  
L E LL ER+A Q +E+ R+ +L+ + L R +L S A R L L  
Sbjct: 358 LTOEVLLLRKVASVESQOEISGRROQLLNLLEG--LVDRSRNLQALQAEOLYSYL 415

Report for DKF2phtes3 7d17.2

[illegible]

[illegible]

Prosite for DKFZphtes3\_7d17.2

PS000001	354-558	ASN GLYCOSYLATION	POC000001
PS000002	315-319	ASN GLYCOSYLATION	POC000002
PS000003	13-116	PKC PHOSPHO_SITE	POC000003
PS000004	25-29	PKC PHOSPHO_SITE	POC000004
PS000005	163-136	PKC PHOSPHO_SITE	POC000005
PS000006	401-404	PKC PHOSPHO_SITE	POC000006
PS000007	188-192	PKC PHOSPHO_SITE	POC000007
PS000008	28-32	PKC PHOSPHO_SITE	POC000008
PS000009	286-290	PKC PHOSPHO_SITE	POC000009
PS000010	295-299	PKC PHOSPHO_SITE	POC000010
PS000011	295-299	PKC PHOSPHO_SITE	POC000011
PS000012	317-321	PKC PHOSPHO_SITE	POC000012
PS000013	317-321	PKC PHOSPHO_SITE	POC000013
PS000014	336-340	PKC PHOSPHO_SITE	POC000014
PS000015	345-349	PKC PHOSPHO_SITE	POC000015
PS000016	372-376	PKC PHOSPHO_SITE	POC000016
PS000017	427-431	PKC PHOSPHO_SITE	POC000017
PS000018	447-451	PKC PHOSPHO_SITE	POC000018
PS000019	502-506	PKC PHOSPHO_SITE	POC000019
PS000020	525-526	PKC PHOSPHO_SITE	POC000020
PS000021	597-601	PKC PHOSPHO_SITE	POC000021
PS000022	25-31	NR1ST1	POC000022
PS000023	201-203	NR1ST1	POC000023

## Plan for DKF2phtes3 7d17.2

HMM_NAME	TNFR/NGFR cysteine-rich region					
HMM	*CpeGtYtDWNHvpqClpCtrCePENGQYMvgPCTwTQTVC*	*	Ce	PEN	GQM	vpgPCTWTQNTVC
	C + ++ N+ ++    ++    +++ ++ ++ ++ VC					
Query	274 CESNOPYG-NT-RITFEEDVDS--TLIDSSSHDEWLDAVC					310

DNF2phtes3.7j3

group: cell cycle

DNF2phtes3.7j3.2 encodes a novel 628 amino acid putative protein kinase, which is related to the C-TAK1/Cdc25C associated protein kinase.

Cdc25C is a protein kinase that controls entry into mitosis by dephosphorylation of Cdc2. Cdc25C function is regulated by phosphorylation, too. Serine 216 phosphorylation of Cdc25C mediates the binding of 14-3-3 protein to Cdc25C. C-TAK1 (Cdc twenty-five C associated protein kinase) phosphorylates Cdc25C on serine 216 in vitro. The new protein is closely related to C-Tak1 and therefore should be involved in cell-cycle regulation, too.

The new protein can find application in modulating/blocking the cell cycle.

strong similarity to serine/threonine-specific protein kinases

complete cDNA, complete cds, potential start at bp 128, few EST hits

Sequenced by BHF2

Locus: unknown

Insert length: 3443 bp  
Poly A stretch at pos. 3399, polyadenylation signal at pos. 3376

```
1  GTGCTTACT  CGCGGCTCTG  GTACTGCTGT  GGCCTCCCGT  CTTGGTGGGG
51  GACCTTGTCG  CGCGGCTCTG  GCCTTCCCGG  CACAGCTTAC  TGATTCCECT
101  GCGGCGCTTG  CTCACCTGCT  GCTGCGCATG  GAGTGGCTGG  TTTTGGGCGG
151  GCCTTCCGCG  CCACTGCTCT  CGGCGCGAGA  GCTAGCCCGG  CGCGTGGCGG
201  AAGGCGTGAT  CAGTGGCGCC  AAGGCTCTTA  TGAGAAGGCA  GGGGTGGAAG
251  CGGACACACC  ACAGGACAAA  CTTGCGGCAC  CGCTACGAGT  TCTTGAGAGC
301  CTTGGGCAAA  GGCACCTACG  GGAAGGTGAA  GAAAGCGCGG  GAGAGCTCGG
351  GCGGCTGCTG  GGCATGAGAG  TCAATCCGGA  AGGACAAAT  CAAGATGAG
401  CAAGATCTGA  TGCACATACG  GAGGGAGATT  GAGATCATGT  CATCACTCAA
451  CCACGCTCAC  ATCAATTGCA  TCCATGAAGT  GTTTGAGAAC  AGCAGCAAGA
501  TCTGTATGCT  CATGAGATAT  GCGAGCGCGG  GCGACCTTTA  TGACTACATC
551  AGCGAGCGCG  ACCAGCTCAG  TGAGCGGCA  GTTAGGGAAT  TCTTCCGGCA
601  GATCGTCTCT  GCGGTGCACT  ATTGCAATCA  GAACAGAGTT  GTCCACCGAG
651  ATCTCAAGCT  GGAGAACATC  CTCCTTGATG  CCAATGGGAA  TATCAAGATT
701  GCTGACTTGG  GCTCTTCCAA  CTTGTACAT  CAAGGGAAGT  TCTTGAGAGC
751  ATTCGTGCGG  AGCGGCTCT  ATGCTCGCG  AGAGATTGTC  AATGGGAAGC
801  CCTACACAGG  CCGAGAGGTG  GACAGCTGCT  CCTGGGTGT  TCTCTCTTAC
851  ATCTGCTGTC  ATGCGACCAT  GCCTTTGAT  GGCATGACC  ATAGATCTCT
901  ACTGAAACAG  ATCAGCAGCG  GGGCTACGCG  GGAGCCAGCT  AAACCTCTCT
951  ATGCTGTGGG  CCGTATCCGG  TGCTTGTGA  TGGTGAACCC  CACCGCGCGG
1001  GCGACGCTGG  AGGATGTGGG  GAGTCACTGG  TGGGTCAACT  GGGGCTACCG
1051  CACCGGAGTG  GAGAGCGAGG  AGGCTCCGCA  TGAGGGTGGG  CAGCTTGGCA
1101  GTGACTCTGG  CGCGGCTTCC  ATGGCTGACT  GCGTCCGGCG  TTCTCCCGCG
1151  CCGCTCTCTG  AGAATGGGCG  CAAAGGTGCG  AGCTTCTTCA  AGCAGCATCG
1201  ACTGTGTGGG  GGAAGGACCA  CCGCTGGCT  GAGGCGGAG  CATTCCTCA
1251  AGAATCTCCG  CAAAGGAAT  GACATGGCC  AGTCTCTCCA  CAGTGACACG
1301  GGTATGACA  CTGCGCATCG  CCGTGGCAAG  AGCAAGCTCA  AGCTGCCAAA
1351  GCGGATTCTC  AAGAGAGAGG  TGTCAAGCTC  TCGCAAGGG  GTACAGAGG
1401  ACCCTCCGGA  GCTCAGCCCA  ATCCCTGCGA  GCGCAGGCGA  GGTGCGCGCG
1451  CTGCTCCCCA  AGAAGGGCAT  TCTCAAGAG  CCGCGACAGC  CGAGTCTGG
1501  CTAGTACTTC  TCTCGGAGC  CAGTGAATC  TGGGAGCTTC  TTGGAGGAG
1551  GCGAGCTGTT  TGTGAGTGGG  GATCCCAAG  AGCAGAGGCC  TCGCAGAGCT
1601  TCAGGGGTGG  TCTCTCATCG  CAAAGGCATC  CTCAGACTCA  ATGGCAAGTT
1651  CTGCGAGACA  GCTTGGAGCG  TCGCGGCGCC  CACCACTTTC  GCTCTCTGG
1701  ATCAACTGCG  CCGACTGCG  CCGCTGGCG  GCGGACAGCC  ACCTCTAGGG
1751  GCTGTAGCG  AGGACAGCAT  CTTCTCTCT  GAGTCTTTTG  ACCAGCTGGA
1801  CTGGCTGAA  CGGCTCCGAG  AGCCGCACT  GCGGGGCTGT  GTGTCTGTGG
1851  ACAAGCTCAC  GCGGCTTGA  GAGCGGCTCT  CAGAGGGGCG  TGAAGCTGTC
1901  CTGAGGCGCT  GCGGCGAGGA  TCTTTGGGG  CAGAGCTGCT  TTTCCCTGAC
1951  AGACTGCCAG  GAGGTGACAG  CGACTACCG  ACAGGACTGG  AGGCTCTGCT
2001  CAAAGCTCAC  CTGAGTGGAG  TAGGCAATCG  CCGAGCGCGG  TCAGGCTCTC
2051  AGATGAGCT  GCTTGCAGC  CAGAGGAGGA  TCTCTCTTC  CCGACTCTCC
2101  AGAAGCTGCA  TCCAGGCTCA  GAGGCTGAG  AGGTTTGA  GTGAGCGCT
2151  GAGAGGGCT  GATATGGGA  AGTAGGCAAA  TGAATGCGC  CAAAGGTTCA
2201  GTTCTGTCT  TCAAGCTGCG  TGAGAGAGA  GATACATAA  GAGAGGGGA
2251  CGGGAATGCC  CGCGACAGAG  TCCATATCG  CTGTTTCTTG  TGTACATGGG
2301  GGGGCGACAG  AGACTGGAA  AGAGACTCT  CCGAGGCGCC  ATCTCTGCA
2351  TCCATGAAT  ACTCTGACA  CATGTGCTCT  TCAAGAGACA  GCTCTTCCG
2401  TACTCATTC  CTGCGAAGT  GGGGCGAGC  CTCTTTACAC  ACACATTCGC
2451  GTTCTTACCA  ACCACGAGAA  CTGATGGTG  GCACCCCTAA  TGTGATGAG
2501  GCATCTCTGG  AATGGTCTGG  AGTAACGCTT  GTTATTTTAT  ATTTTATTTT
```

WO 01/12659

PCT/IB00/01496

```

2551 TTAATATTTT ATTTATTTTT TTGAGACGGA GTTTCGGCTT TGGTGGCCAG
2601 GCTAGAGTGC AATGGCGGGA TCTCAGTCA CCTCAACCTC GGCCTCCGGG
2651 GTTCAAGCGA TTCTCCTGCC TCAGCCTCCC TAGTACGTGG GATTACAGGC
2701 GCGCGCCACC ATGCGCGGCT AATTTTGTAT TTTTAGTAGA GACAGGOTTT
2751 CTCAGTGTTC GTCAGGCTGG TCTCAACCTC CGACCTTCAG GTGATCAGC
2801 CACCTCGGCC TCCCAAAGTG CTGGGATTAC AGCGGTGAGC CACCGCGCCC
2851 CACCTAACCC TTCTTATTTT AGCCTAGGAG TAGAGAGACA CAATCTCTGT
2901 TCTTCAATG GTTCTCTTCC CTCTTCATC CTCGAAACCT GCGCTAGGCC
2951 TCTGAAGTTT GCTGCTGTGA ATCTGAAGA CTGAAAGC CTCCGCTGCG
3001 TGTGTGACT TCATCTCAAG GCGGCCAGCC TCTCTGGAG TCCACTTGG
3051 ACCTCAGTGA CTCAGAACTT CTGCTCTTAA GCTGCTTAA ACTCCAGACT
3101 ATGATCTCTT TCTCTAGGCC TTGAGGACTC TAGAATGTC ATATTATTTT
3151 TTATGTTCTT GCGTTTGTCT TTAGGAGAAA GTGAATCTTG CTGTTTTCAA
3201 TAATGTGAAT GCTATGTTCT GGGAAATCCC ACTATGACAT CTAAGTTTTC
3251 TGTAGAGAGA GATATTTTTC CAATATTCCT CAGCTCTTCC CAGAACCCCC
3301 CACACTCCAC TCCCACTCT TGAATCTCTT TACCTAATGG TCTCTACCTA
3351 ATGGACCTCC GTGGCCAAAA AGTACCATTAA AAACCGAGAA GGTGATTGGA
3401 AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAA

```

# BLAST Results

No BLAST result

## Medline entries

98202387:  
C-TAK1 protein kinase phosphorylates human Cdc25C on serine 216 and promotes 14-3-3 protein binding.

## Peptide information for frame 2

ORF from 128 bp to 2011 bp; peptide length: 628  
Category: strong similarity to known protein

```

1 MESLVFARRS GPTPSAAELA RPLAEGLIKS KPPLMKQAV KRHHKHNLK
51 HRYEFLLETIG KGTGKVRKA RESSGRLVAI KSIRKDKND DDLAMHRRK
101 IELHSLAHV HIALHEVTE KSRKIVVME YASRGDTOT ISEKQLSER
151 EAKHFTFROIV SAVHYCHQNR VVHRDLKLEH ILDANGNIK IADFGLSMLY
201 HOGKFLQTFC GSPLVASPEI VNGKPTTQPE VDSHSLGVLL YILVHGTMPF
251 GGDHRIKLVK QISNGAYREP PFPSDAGCLI RMLAKWVPTA RATLEDVASH
301 WVMWGVATK VGEQAPHEG GPCSDSARA SKADWLRRSS RPLENGAKV
351 CSFFKQHAPO GGSTTPGLER QHSLKSRKE NMAQSLHSD TADDTAHRNG
401 KSKLRLPKGI LKKNVASAL GVQEDPFELS PIPASPGDAA PLLPKGILK
451 EPNQRESQIV SSEPSSESG LIDAGDVFS GDFKQKFPQ ASGLLLHRRG
501 ILKNGKFSQ TALELAAPT FGLDELAPP RFLARASRPS GAVSEDSILS
551 SESTDLQLDP ERLPERPLAG CYSVONLTGL EEPSPSGRPS CLRWRRQDFL
601 GDSCESLTDC QEVATVYRGA LRWCKSLT

```

## BLASTP hits

No BLASTP hits available

Alert BLASTP hits for ORFsphtes3\_7j3, frame 2

No Alert BLASTP hits found

Peptide information for ORFsphtes3\_7j3, frame 2

## Report for ORFsphtes3\_7j3.2

```

[LENGTH] 628
[FW] 69612.39
[pI] 9.01
[ROMOL] TREMBL:AB011109.1 gene: "KIAA0537"; product: "KIAA0537 protein"; Homo sapiens
[UNCAT] KIAA0537 protein, complete cds. 1e-152
[UNCAT] 01.05.04 regulation of carbohydrate utilization [S. cerevisiae, YDR477w]
5e-66
[UNCAT] 11.01 stress response [S. cerevisiae, YDR477w] 5e-66

```

[FUNCAT] 30.03 organization of cytoplasm [S. cerevisiae, YDR477w] 5e-66  
[FUNCAT] 98 classification not yet clear-cut [S. cerevisiae, YLR096w] 6e-54  
[FUNCAT] 30.02 organization of plasma membrane [S. cerevisiae, YLR096w] 6e-54  
[FUNCAT] 03.04 budding, cell polarity and filament formation [S. cerevisiae, YDR507c]  
8e-52  
[FUNCAT] 03.25 cytokinesis [S. cerevisiae, YDR507c] 8e-52  
[FUNCAT] 03.22 cell cycle control and mitosis [S. cerevisiae, YKL101w] 9e-51  
[FUNCAT] 30.10 nuclear organization [S. cerevisiae, YKL101w] 9e-51  
[FUNCAT] 99 unclassified proteins [S. cerevisiae, YPL141c] 1e-45  
[FUNCAT] 10.99 other signal-transduction activities [S. cerevisiae, YPL153c] 6e-44  
[FUNCAT] 03.22.01 cell cycle checkpoint proteins [S. cerevisiae, YPL153c] 6e-44  
[FUNCAT] 11.04 dna repair (direct repair, base excision repair and nucleotide excision  
repair) [S. cerevisiae, YPL153c] 6e-44  
[FUNCAT] 03.19 recombination and dna repair [S. cerevisiae, YPL153c] 6e-44  
[FUNCAT] 03.16 dna synthesis and replication [S. cerevisiae, YHR001c] 2e-42  
[FUNCAT] 10.02.11 key kinases [S. cerevisiae, YBL105c] 3e-34  
[FUNCAT] 04.05.01.04 transcriptional control [S. cerevisiae, YKL139w CTK1 - carboxy-  
terminal domain] 2e-28  
[FUNCAT] 03.01 cell growth [S. cerevisiae, YFR014c] 4e-28  
[FUNCAT] 03.10 sporulation and germination [S. cerevisiae, YGL180w] 2e-26  
[FUNCAT] 06.13.04 lysosomal and vacuolar degradation [S. cerevisiae, YGL180w] 2e-26  
[FUNCAT] 08.13 vacuolar transport [S. cerevisiae, YGL180w] 2e-26  
[FUNCAT] 04.99 other transcription activities [S. cerevisiae, YER129w] 4e-26  
[FUNCAT] 02.19 metabolism of energy reserves (glycogen, trehalose) [S. cerevisiae,  
YPL031c] 5e-24  
[FUNCAT] 01.04.04 regulation of phosphate utilisation [S. cerevisiae, YPL031c]  
5e-24  
[FUNCAT] 03.07 pheromone response, mating-type determination, sex-specific proteins  
[S. cerevisiae, YHL007c] 6e-24  
[FUNCAT] 10.05.11 key kinases [S. cerevisiae, YHL007c] 6e-24  
[FUNCAT] 09.01 biogenesis of cell wall [S. cerevisiae, YHR031c] 1e-22  
[FUNCAT] 10.03.11 key kinases [S. cerevisiae, YHR031c] 1e-22  
[FUNCAT] 03.13 meiosis [S. cerevisiae, YDR523c] 8e-22  
[FUNCAT] 04.05.01.01 general transcription activities [S. cerevisiae, YDL108w]  
6e-21  
[FUNCAT] 06.07 protein modification (glycosylation, acylation, myristylation,  
palmitoylation and processing) [S. cerevisiae, YPL031c] 6e-21  
[FUNCAT] 10.05.09 regulation of p-protein activity [S. cerevisiae, YBL016w] 7e-19  
[FUNCAT] 10.04.11 key kinases [S. cerevisiae, YDL159w] 3e-18  
[FUNCAT] 01.02.04 regulation of nitrogen and sulphur utilization [S. cerevisiae,  
YNL183c] 1e-17  
[FUNCAT] 08.99 other intracellular-transport activities [S. cerevisiae, YNL183c]  
1e-17  
[FUNCAT] 05.07 translational control [S. cerevisiae, YDR283c] 2e-17  
[FUNCAT] 09.04 biogenesis of cytoskeleton [S. cerevisiae, YML020c] 4e-16  
[FUNCAT] 04.03.99 other trna-transcription activities [S. cerevisiae, YOR061w]  
1e-15  
[FUNCAT] 10.04.99 other nutritional-response activities [S. cerevisiae, YJR059w]  
5e-15  
[FUNCAT] c energy conversion [M. genitalium, MG109] 3e-12  
[FUNCAT] 30.09 organisation of intracellular transport vesicles [S. cerevisiae,  
YBR097w] 2e-08  
[FUNCAT] 08.07 vesicular transport (golgi network, etc.) [S. cerevisiae, YBR097w]  
2e-08  
[FUNCAT] 06.04 protein targeting, sorting and translocation [S. cerevisiae, YBR097w]  
2e-08  
[FUNCAT] 30.08 organisation of golgi [S. cerevisiae, YBR097w] 2e-08  
[FUNCAT] 30.07 organisation of endoplasmatic reticulum [S. cerevisiae, YHR079c]  
8e-05  
[FUNCAT] 01.06.10 regulation of lipid, fatty-acid and sterol biosynthesis [S.  
cerevisiae, YHR079c] 8e-05  
[BLOCKS] BL00479C Phorbol esters / diacylglycerol binding domain proteins  
[BLOCKS] BL00239B Receptor tyrosine kinase class II proteins  
[BLOCKS] BL00107A Protein kinases ATP-binding region proteins  
[SCOP] digol\_5.1.1.1.9 MAP kinase Erk2 (rat Rattus norvegicus) 1e-77  
[SCOP] dlwfc\_5.1.1.1.8 MAP kinase p38 (human (Homo sapiens)) 4e-68  
[SCOP] dlkaa\_5.1.1.1.7 (1-350) Twitchin, kinase domain (Caenorhabditis) 2e-85  
[SCOP] dlkba\_5.1.1.1.6 Twitchin, kinase domain (california sea har) 1e-80  
[SCOP] dlphk\_5.1.1.1.5 gamma-subunit of glycogen phosphorylase kinase 2e-76  
[SCOP] dlirk\_5.1.1.2.4 insulin receptor (human (Homo sapiens)) 1e-69  
[SCOP] dlapne\_5.1.1.1.4 cAMP-dependent PK, catalytic subunit (mouse (Mus)) 1e-84  
[SCOP] dlfgka\_5.1.1.2.3 Fibroblast growth factor receptor 1 (human (Homo sapiens)) 1e-68  
[SCOP] dlydra\_5.1.1.1.3 cAMP-dependent PK, catalytic subunit (bovine (Bos)) 9e-85  
[SCOP] dlfaa\_5.1.1.1.2 (168-437) c-src tyrosine kinase (human (Homo sapiens)) 1e-69  
[SCOP] dldka\_5.1.1.1.2 cAMP-dependent PK, catalytic subunit (pig (Sus)) 1e-85  
[SCOP] dlhka\_5.1.1.2.1 (167-437) Haemopoietic cell kinase Hck (human (Homo sapiens)) 5e-66  
[SCOP] dlkaa\_5.1.1.1.11 Casein kinase-1, CK1 (Schizosaccharomyces pombe) 9e-47  
[SCOP] dljaa\_5.1.1.1.10 Casein kinase-1, CK1 (human (Homo sapiens)) 1e-75  
[SCOP] dlkja\_5.1.1.1.10 Casein kinase-1, CK1 (rat (Rattus norvegicus)) 5e-54  
[EC] 2.7.1.38 Phosphorylase kinase 1e-36  
[EC] 2.7.1.123 G2b/calmodulin-dependent protein kinase 4e-40



[EC] 2.7.1.128 [Acetyl-CoA carboxylase] kinase 1e-61  
 [EC] 2.7.1.117 Myosin-light-chain kinase 2e-40  
 [EC] 2.7.1.109 [Hydroxymethylglutaryl-CoA reductase(NADPH)] kinase 1e-61  
 [EC] 2.7.1.37 Protein kinase 7e-42  
 [PIRKEW] phosphotransferase 6e-66  
 [PIRKEW] nucleus 1e-64  
 [PIRKEW] calcium 7e-35  
 [PIRKEW] duplication 1e-38  
 [PIRKEW] tandem repeat 4e-39  
 [PIRKEW] phorbol ester binding 1e-38  
 [PIRKEW] zinc 1e-38  
 [PIRKEW] cell cycle control 1e-42  
 [PIRKEW] serine/threonine-specific protein kinase 8e-68  
 [PIRKEW] oncogene 1e-40  
 [PIRKEW] phospholipid binding 1e-38  
 [PIRKEW] autophosphorylation 1e-64  
 [PIRKEW] brain 1e-40  
 [PIRKEW] heterodimer 2e-36  
 [PIRKEW] mitosis 7e-42  
 [PIRKEW] polymer 1e-35  
 [PIRKEW] magnesium 6e-66  
 [PIRKEW] ATP 8e-68  
 [PIRKEW] polyprotein 1e-40  
 [PIRKEW] phosphoprotein 1e-64  
 [PIRKEW] apoptosis 4e-39  
 [PIRKEW] glycoprotein 7e-42  
 [PIRKEW] leucine zipper 3e-35  
 [PIRKEW] skeletal muscle 7e-35  
 [PIRKEW] protein kinase 5e-41  
 [PIRKEW] cAMP binding 3e-38  
 [PIRKEW] testis 9e-36  
 [PIRKEW] purine nucleotide binding 2e-49  
 [PIRKEW] calcium binding 8e-39  
 [PIRKEW] alternative splicing 3e-37  
 [PIRKEW] P-loop 2e-49  
 [PIRKEW] lipoprotein 2e-33  
 [PIRKEW] segmentation 1e-33  
 [PIRKEW] core protein 1e-40  
 [PIRKEW] muscle 7e-35  
 [PIRKEW] myristylation 2e-33  
 [PIRKEW] EF hand 8e-39  
 [PIRKEW] cell division 2e-40  
 [PIRKEW] calmodulin binding 4e-40  
 [SUPFAM] ribosomal protein S6 kinase II 5e-36  
 [SUPFAM] fibronectin type III repeat homology 3e-33  
 [SUPFAM] immunoglobulin homology 3e-33  
 [SUPFAM] calcium-dependent protein kinase 8e-39  
 [SUPFAM] AMP-activated protein kinase 6e-66  
 [SUPFAM] protein kinase act 3e-42  
 [SUPFAM] protein kinase SPK1 1e-42  
 [SUPFAM] unassigned Ser/Thr or Tyr-specific protein kinases 8e-68  
 [SUPFAM] Ca2+/calmodulin-dependent protein kinase 3e-37  
 [SUPFAM] calmodulin repeat homology 8e-39  
 [SUPFAM] cAMP receptor protein cyclic nucleotide-binding domain homology 6e-33  
 [SUPFAM] protein kinase C zeta 1e-36  
 [SUPFAM] Dictyostelium cAMP-dependent protein kinase catalytic chain 1e-34  
 [SUPFAM] death-associated protein kinase 4e-39  
 [SUPFAM] plectrin repeat homology 3e-42  
 [SUPFAM] ankyrin repeat homology 4e-39  
 [SUPFAM] protein kinase homology 8e-68  
 [SUPFAM] Ca2+/calmodulin-dependent protein kinase II 8e-41  
 [SUPFAM] protein kinase C zinc-binding repeat homology 1e-38  
 [SUPFAM] twitchin 3e-33  
 [SUPFAM] protein kinase C delta 1e-38  
 [SUPFAM] cAMP-dependent protein kinase 6e-33  
 [SUPFAM] protein kinase cdrl 7e-42  
 [SUPFAM] protein kinase C C2 region homology 3e-37  
 [SUPFAM] protein kinase C alpha 3e-37  
 [SUPFAM] yeast protein kinase C 5e-36  
 [SUPFAM] kinase-related transforming protein 1e-41  
 [SUPFAM] kinase interaction domain homology 1e-42  
 [SUPFAM] Gsp-act polyprotein 1e-40  
 [SUPFAM] Ca2+/calmodulin-dependent protein kinase I 4e-40  
 [SUPFAM] protein kinase C mu 4e-33  
 [PROSITE] PROSITE\_KINASE\_ATP 2  
 [PROSITE] RGD 1  
 [PROSITE] MYRISTYL 4  
 [PROSITE] CAMP\_PHOSPHO\_SITE 3  
 [PROSITE] CK2\_PHOSPHO\_SITE 13  
 [PROSITE] TYR\_PHOSPHO\_SITE 2  
 [PROSITE] PKC\_PHOSPHO\_SITE 12

```
(PROSITE)    ASN_GLYCOSYLATION      2
(PROSITE)    PROTEIN_KINASE_ST      1
(PFAM)       Eukaryotic protein kinase domain
(KW)         All_Alpha
(KW)         30
(KW)         LOW_COMPLEXITY      10.51 %

SEQ    MESLVFARRSGPTPSAAELARPLAEGLIKSPKPKKQAVKHHKHNLRHYEFLETIG
SEG    .....XXXXXXXXXXXXX.....
lc1pE  .....HKKHHHHHHHHHHHCCCCCCCC--GGGEEEEEEEE

SEQ    KGYGVKVKARESSGRLVAKSIKDKIKDEQDLMIIRREIETMSSLNHPHIIATIEVFE
SEG    .....
lc1pE  CTTTEEEEEETTTTEKEEEELKHHHHHHHHHHHHHHHHHHHHHCCCTTTTCCCEEEEEEE

SEQ    NSSKIVIVNEYASRGDLVDYISENQQLSEREARHFFQIVSAVHYCHQNRVVRDLALEH
SEG    .....
lc1pE  ETTTEEEEEETTTTBNHHHHHHHCCCCCHHHHHHHHHHHHHHHHHHHHCCCECCCCGGG

SEQ    ILLDANGHRIADPGLSHLYHGFQGTFCGSLYASFEIVNGKPYTGPEVDSMSLGVL
SEG    .....
lc1pE  EETTTTCEEECCCTTTTET-TT-BCCCCCGGGCHHHHHCCCCB-HHHHHHHHHHHH

SEQ    YILVHGTFNFDGHHKILVKQISNGAYREPPPSDACGLRWLMVNPTRATLEDVASH
SEG    .....
lc1pE  HHHHCCCTTTTTHHHHHHHHHHCCCCCTTCHHHHHHHHHHTTTTGGGTTTNNHHHHC

SEQ    WVNNGYATRVGEQAPREGGHPGSDSARASADNLARSSRPLENGARVCSFFQKAPG
SEG    .....
lc1pE  GG.....

SEQ    GGSITPGLERQHSLLKSRKENDMAQSLHSOTADDTAHRPGKSHLKPGLIKKKVSAE
SEG    .....
lc1pE  .....

SEQ    GVQEDPFELSPFASPGQAAPLLPKYGLKTPQKRESGVYSSPEPSKGLLDAGDVVS
SEG    .....XXXXXXXXXXXXX.....XXXXXXXXXXXXX.....
lc1pE  .....

SEQ    GDFKEQKPPQASGLLLKHKGILKMGKFSQALELAAPTTFGLSDELAAPPRLARASRPS
SEG    .....XXXXXXXXXXXXX.....
lc1pE  .....

SEQ    GAVSEDSILSSESPDQLDLPALPEPPLAGCVSDWLTGLEPPSEGPQSLARWRQDPL
SEG    .....XXXXXXXXXXXXX.....
lc1pE  .....

SEQ    GDSCFSLTDCQEVATATYQALRVCSKLT
SEG    .....
lc1pE  .....
```

## Prosite for DKFiphtes3\_tj3.2

PS00001	121->125	ASN_GLYCOSYLATION	PD0C00001
PS00001	576->580	ASN_GLYCOSYLATION	PD0C00001
PS00004	290->294	CAMP_PHOSPHO_SITE	PD0C00004
PS00004	337->341	CAMP_PHOSPHO_SITE	PD0C00004
PS00004	413->417	CAMP_PHOSPHO_SITE	PD0C00004
PS00005	30->33	PKC_PHOSPHO_SITE	PD0C00005
PS00005	74->77	PKC_PHOSPHO_SITE	PD0C00005
PS00005	82->85	PKC_PHOSPHO_SITE	PD0C00005
PS00005	122->125	PKC_PHOSPHO_SITE	PD0C00005
PS00005	142->145	PKC_PHOSPHO_SITE	PD0C00005
PS00005	148->151	PKC_PHOSPHO_SITE	PD0C00005
PS00005	289->292	PKC_PHOSPHO_SITE	PD0C00005
PS00005	327->330	PKC_PHOSPHO_SITE	PD0C00005
PS00005	339->342	PKC_PHOSPHO_SITE	PD0C00005
PS00005	373->376	PKC_PHOSPHO_SITE	PD0C00005
PS00005	377->380	PKC_PHOSPHO_SITE	PD0C00005
PS00005	616->619	PKC_PHOSPHO_SITE	PD0C00005
PS00006	13->19	CK2_PHOSPHO_SITE	PD0C00006
PS00006	133->137	CK2_PHOSPHO_SITE	PD0C00006
PS00006	148->152	CK2_PHOSPHO_SITE	PD0C00006
PS00006	227->231	CK2_PHOSPHO_SITE	PD0C00006
PS00006	293->297	CK2_PHOSPHO_SITE	PD0C00006
PS00006	331->335	CK2_PHOSPHO_SITE	PD0C00006
PS00006	377->381	CK2_PHOSPHO_SITE	PD0C00006
PS00006	391->395	CK2_PHOSPHO_SITE	PD0C00006

PS00006	461->465	CK2_PHOSPHO_SITE	PDOC00006
PS00006	511->515	CK2_PHOSPHO_SITE	PDOC00006
PS00006	523->527	CK2_PHOSPHO_SITE	PDOC00006
PS00006	576->582	CK2_PHOSPHO_SITE	PDOC00006
PS00006	606->610	CK2_PHOSPHO_SITE	PDOC00006
PS00007	453->460	TYR_PHOSPHO_SITE	PDOC00007
PS00007	453->461	TYR_PHOSPHO_SITE	PDOC00007
PS00008	320->326	MYRISTYL	PDOC00008
PS00008	324->330	MYRISTYL	PDOC00008
PS00008	347->353	MYRISTYL	PDOC00008
PS00008	360->366	MYRISTYL	PDOC00008
PS00016	134->137	RGD	PDOC00016
PS00107	59->82	PROTEIN_KINASE_ATP	PDOC00100
PS00107	59->86	PROTEIN_KINASE_ATP	PDOC00100
PS00108	171->184	PROTEIN_KINASE_ST	PDOC00100

## Pfam for DKFZp483\_7j3.2

HMM_NAME	Eukaryotic protein kinase domain
HMM	YelqRlIcGafGLVvVcLWtGcIvAIKlIkKams.....FIREI
Query	53 YEFLETLGKGTGKVKKARSSGRVAKSIRKDKIKDEQDLWHRIRREI 101
HMM	qINRcLhNHPWIRFYDwFeddDhIVIMNEYMeGDLFDYIrnqPMSw
Query	102 EIMSSLWHPHIIAINEVFE-NSSKIVIVNEYASRGDLVDYISERQQLSER 150
HMM	eTcflNyQILrGMeYLHSMqIIRDLRPNILIdNgqTKicDFGLARqM
Query	151 EARHFFRQIVSAVYCHQNRVVRDLKLENILLDANGHRIADFGLSHLY 200
HMM	nnYzmtLfcGTWYMAPEVIlmg.nyYtLlVDMMSFGCILMENDNTGep
Query	201 HQGFLQTTCGSPLTA-SPEI-VNGRPTTCPEVDMSLGVLTLILVHCTH 248
HMM	PfYddnNemImrIlq:frpDwpcSeElYdFhwCWnyDPeKRPtF:QI
Query	249 PFDGNDKILVKQISNGATREPPKPSD-ACGLIRLLMVPNTERATLIDV 297
HMM	LhNPMF*
HMM	H W*
Query	298 ASHMMV 303

WO 01/12659

PCT/IB00/01496

DKFSphtes3\_7j8

group: testes derived

DKFSphtes3\_7j8 encodes a novel 410 amino acid protein nearly identical to human WUGSC:H\_DJ1159004.1.

The novel protein contains an additional C-terminal domain, which is not present in WUGSC:H\_DJ1159004.1.  
No informative BLAST results; No predictive prosite, pfam or SCOP motif.

The new protein can find application in studying the expression profile of testis-specific genes.

WUGSC:H\_DJ1159004.1 similarity to YBL104p

verifies and extends the genmodel WUGSC:H\_DJ1159004.1  
similarity to S.cerevisiae YBL104p

Sequenced by BNF2

Locus: /map="7p21-p22"

Insert length: 3353 bp  
Poly A stretch at pos. 3231, no polyadenylation signal found

```

1 GCAGAAATATG TTGTATTGT GGCATAGTC ATATTACAC TATCATAAAA
51 TTATGGCCGA GAAGTAAAT ATTCTAAATG TGTCACATA GTTCTCTGTA
101 AAATGACATT ATTTCCAAA TATATTTTGA AATAAACAA TATAAAAAATG
151 TTTCCTGTTT TTAGCAATGG TGGAAACAG CAGACATAAT TGGAGTGGGT
201 TGATATAGCA AGTGTATATT CAAGATTTAA ATGAAGAGAG AATCTTAGCT
251 TTACAGCTTT GTGGTGGAT AAAGAAGGA ACGGATGTAG ACOTGGGGCC
301 ATTTTGAAC TCCCTGTGAC AAGAAGGGA ATGGCAAGA GCTGCTGCTG
351 TGGCATTTCT CACTTGTGAT ATTGGCGAG CAATCAAT CTGAATGAA
401 GGGGCATCTT CTGAAAAGG AGATCTGAAT CTCAATGGG TAGCAATGGC
451 TTTATCGGGT TATACGGATG AGAAGAACTG CCTTTGGAGA GAAATGTGTA
501 GCACACTGCG ATACAGCTA AATACCCCT ATTTGTGTG CATTTTGGCA
551 TTTCTGACAA GTGAACAGG ATCTTAGCAT GGAGTTTGT ATGAAGACAA
601 AGTTGCAGTA CGTCACAGAG TGGCATTGCG TTGTAAATTC CTTAGTGATA
651 CTCAGTAAAT TACATACAT CAAGATTGTA CCATGAAAT GAAGAGGCT
701 GCAAAATTGG AAGGATTTT GCTTACAGGC CTTACTAAG ATGGAGTGA
751 CTTAATGGAG AGTTATGTT ATAGAAGTGG AGATOTCAA ACAGCAAGTT
801 ACTGTATGTT ACAGGTTTCA CTTTATAGT TTTCTTAAGA TGAAGGGTT
851 CAGTACTGGA TTGCAATTA TAGAAATTA TTAGATGCTT GGAGCTTTTG
901 GCATAAAGCA GCTGAATTTG ATATTCAGG GACTAAGTG GATCCAGATT
951 CCAAGCCTTT AGCACAAGTT TTTGTGAGTT GCAATTTCGT TGGCAAGTCA
1001 ATCTCTACA GCTCTTACG TTGCTCTCAT CAGGCAAGG GTTTTAGTCA
1051 GTATGGTGTG AGTGGCTCAC CAACGAATC TAAAGTCACA AGTTTCTCTG
1101 GCTGTGAAA ACCACTTCTT CGATGTGGCG TTTGTCTCAT TAATATGGGA
1151 ACACCAATTT CTACCTCTCC TGGAGGAAC AAATCAGATG AAAAGTGA
1201 CTTGAGCAGG GACAAAAAAT TAGCCCAATT TAACAAGTGG TTTACATGGT
1251 GTCATAATG CAGGCACGGT GGACATCTGT GACATATGCT TACTTGGTTC
1301 AGGACACATG CAGAGTGGCC TTGTGTGCTG TGCACGTGTA AATATATGCA
1351 GTTGATACCA ACGGGAATC TCGTACTCTG AGAAGCTCTC CAGCATAAAT
1401 ATCTTACCAC CTTAAGAGAA CCTTCAAGT GTGGAGCTTT CTAGTAGGTT
1451 TCCTTATAG CTTGAGACCA TACTTCAGAA CAGGCATTC ATGACTTACC
1501 TGTAAATGGG AATTAATCA TCTATACAG TCAGCAGTTT TGATTTTGA
1551 GTGATTTTGA TATGCTTCAC AGAGACAAT GTGCCAATA TAAACATGGA
1601 AGTATAGACA TGAGTTCTGT TCAGCAGGTT GAAAAGTCTG ATTTAGAAAA
1651 ACTTCTAAG TTTTGGTGA AATATGAC ACTGTAGAG CAGATTTCT
1701 GGAGAGCCA AGAACAGCT TTGAGCTAT ATCTTCAAG CTGAAGCTGG
1751 ATATCTTTCA ATAAATATG TGCACTTTAA AATAAATG ACTAATCTG
1801 TGATTACAG AATAGTTTAA AGTTCAGCT TGCTTAAAT TCTTTCAGAT
1851 TAATTAAATA TTATGATTTT TACTTTTAG AATGACAGC CCGTATGCC
1901 ACACGGAGA ATATTTTAAA TACTGTCTGT TTATATATGT GTCTATGTGT
1951 GTGTGTATAT TTATGTGTGT AGTATAAAT ATGTACTTTT TAAAGAGGCC
2001 TTTTGGCTCC TTGATTTTAA AGATAGCAA TCTTTTGGCA TAACATTATC
2051 GTCTTCTAG AAAAGCCAAG ATGAAGATC TATCTTACAA CTTTTCCTCT
2101 TCAGTAGAGA AAAACATGTA CCAITTCAGG TGAACATACA AAATTTTCAC
2151 TTTTACCTTT TGGCTTCCA ATGCTCTGAT TTGCTTCAA AGGTTTCTCT
2201 CCAATATTA TTTGCACTT ATCTTCATCA CTTGAGACA TTTTACTGCA
2251 TACAAGTCT ATGCAAGATT ATATGTAAT AGCCATTAG TATAATCTAT
2301 GTCACTGTTT CTGTGCTGTC AATTCCTCT CTGATTTGGA ATACCAATCC
2351 TTTTCTTTC CAGGTAGAC TGGCACTGT TGGCAATAA GGCTCACTTC
2401 AGAGACCATT TTAGATGTAA GTTTTAAAT TTAAGTGTTA CTGGGGCTAA
2451 GTCAAGGACT TTATTTAAAA CATTTTFTT TCTCATTTCT ATAGCTAGAT
2501 AGTTGTAAAG GAAATACAAA GAATTTACAA GATGCTTCTG TGTCTCTG

```

WO 01/12659

PCT/IB00/01496

```
2551 CGTATGCAGA GGAAGTGAC TAGGAATTTT GTAGTTGAAG CTGTGTCAT
2601 AAGAGCTTAA TCTTATTTA TAGATTTGG AGAATAAATA CAGAATTTT
2651 AAGAGCTTTC GTATTAGCAG TTTTGCCCTA TAAAACTAA GATTGTGAG
2701 ATTAGTTTGA GGTGTAACTT AATATTAAA ACTGATTAAA ATTATTTT
2751 TACCTTGAGT GTCTGATACA TAAACCCCTT TCTAGGAAA ACATTGGAG
2801 TAGTACATAT TTACTCTAAA TGCTCACCT GCATGACAGT CTTTCAAA
2851 GAAGACATAT GTAAATGAAA TTTTCTTTA AGATTGCTA TTAAGGGTAC
2901 TTTTCCAGC CTTCATTGA GTAAATCTTA ATTGATTCA TTTTATTAAC
2951 ATATACCCCT TACCTTTAT ATTTCAATTG AGTGTTCCT TCAAACTTA
3001 CTGCTTAAA TATGAAGTC AGCTTAACT AATGTGACAC TCATATGAT
3051 TTCAATCTC ATTACTAAA GTAAATGTA AATTATCTC AATAGTTAC
3101 AGCTTTTGA AATACATAT AAAACATGAA TGTAAAGTCT ATATGTAA
3151 ATGCTTATTT GTATCTTAA TATATGAGG TGACATTTT AGATTGTAT
3201 GTATGTGCTA ACCTCTTAA TGTCTTCTT GAAAAAATA AAAAAAATA
3251 AAAAAAATA AAAAAAATA AAAAAAATA AAAAAAATA AAAAAAATA
3301 AAAAAAATA AAAAAAATA AAAAAAATA AAAAAAATA AAAAAAATA
3351 AAA
```

BLAST Results

No BLAST result

Medline entries

No Medline entry

Peptide information for frame 2

ORF from 167 bp to 1396 bp; peptide length: 410  
Category: known protein  
Classification: unclassified

```
1 MVESSRNHWS GLDQSDIOM LNEERILALQ LCGWIKKGTQ DVDPFPLKSL
51 VQEGEMERAA AVALFNLDIR RAIQILNEGA SSEKGDNLN VVAHALSGYT
101 DEKNSLWREN CSTLRQLQWN PYLCVNFAPL TSETGSDGV LYENKVVVRD
151 RVAFACKTLE DYOLWRYIEK LTHWKEACH LDIILLTGLT KQGVLMESY
201 VORTQDVQTA SYCHLQSPFL DVLKDERVOY WIEHYRNLLD AWRFMKRAE
251 FDIRSKLQSP SSKPLAQVIV SCHPCGASIS VSCSAVPHQG RGFSDYGVSG
301 SPFKSKVISC PCRRPLPKPC ALCLINWGPV VSSCPGQTSR DENVOLSKDK
351 RLADFNWMTF WCHNCRHUGH AGHKLWFRD HXCPVSACT CKMQLDTTG
401 HLVPATVOP
```

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKEFphes3\_738, frame 2

PIR:S45391 probable membrane protein YBL104c - yeast (Saccharomyces cerevisiae), N = 2, Score = 446, P = 4.5e-47

TREMBL:AC004982.1 gene: "WUGSC:H\_DJ1159004.1"; Homo sapiens PAC clone DJ1159004 from 7p21-p22, complete sequence., N = 1, Score = 2038, P = 7.6e-211

>TREMBL:AC004982.1 gene: "WUGSC:H\_DJ1159004.1"; Homo sapiens PAC clone DJ1159004 from 7p21-p22, complete sequence.  
Length = 379

HSPs:

Score = 2038 (305.8 bits), Expect = 7.6e-211, P = 7.6e-211  
Identities = 379/379 (100%), Positives = 379/379 (100%)

```
Query: 1 MVESSRNHWSGLDQSDIOMLNEERILALQQLCGWIKKGTQDVDPFPLKSLVQEGEMERAA 60
      MVESSRNHWSGLDQSDIOMLNEERILALQQLCGWIKKGTQDVDPFPLKSLVQEGEMERAA 60
Sbjct: 1 MVESSRNHWSGLDQSDIOMLNEERILALQQLCGWIKKGTQDVDPFPLKSLVQEGEMERAA 60

Query: 61 AVALFNLDIRRAIQILNEGASSEKGDNLNVVAHALSGYTDEKNSLWREKSTLRQLQWN 120
      AVALFNLDIRRAIQILNEGASSEKGDNLNVVAHALSGYTDEKNSLWREKSTLRQLQWN 120
Sbjct: 61 AVALFNLDIRRAIQILNEGASSEKGDNLNVVAHALSGYTDEKNSLWREKSTLRQLQWN 120
```

Report for DKFZphtes3 718.2

```
SRD      NVWSSHHHSLGLOFQSDGHHHEERLALQLQGVF:KXGTVDWGVGFUNSLVQGEDEMARAA
SRD      cccccccccccccccccchhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh
SRD
SRD      AVALLVLLDRAAQLLWZGASGHHGHHLLVVAWALSGVYOTEKSSLRPDCSTALZLQVH
SRD      hhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh
SRD
SRD      PLYLWCFALFTSTGSDGVLYEHVAVQDPAFACFLHSTOYHHHLLIETLWKEHGAQIG
SRD      cccccccccccccccccchhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh
SRD
SRD      LEEILLTGLGVHMLSHVYDGTQWYATSYLQSGFVLDKDEYVWLYEHMLLD
SRD      cccccccccccccccccchhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh
SRD
SRD      AMRTHRAEETLHSHSLDFFSLQAVFYSCFPGKSIYSGSYAHGAGFSYQYSG
SRD      hhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh
SRD
SRD      SPTKASVTCPCPRFLNPLCALILMGTVPSSGFGTSSDEZVLSKOKLQAFNWFTE
SRD      ccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc
SRD
SRD      MCNCHPGHGHGHHLSHVRMACDFYSACTKQGLQTTGNLPACTVQV
SRD      ccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc
```

(No Pfam data available for DKFZphtes3 718.2)

WO 01/12659

PCT/IB00/01496

DKFphtes3\_7p10  
-----

group: Cell Cycle

DKFphtes3\_7p10.1 encodes a novel 422 amino acid putative protein, which is closely related to the *Xenopus laevis* XPMC2 protein.

In fission yeast the kinases Wee1 and Mik1 control that initiation of mitosis starts after completion of DNA synthesis. Yeast in which both Wee1 and Mik1 kinases are defective exhibit a mitotic catastrophe phenotype. XPMC2 of *Xenopus* rescues several different yeast mitotic catastrophe mutants defective in Wee1/Mik1 kinase function. The XPMC2 protein is localised in the nucleus in *Xenopus* oocytes. The new protein is the human orthologue of this gene.

The new protein can find application in modulating/blocking the cell cycle.

strong similarity to XPMC2 protein

complete cDNA, complete cds, EST hits

Sequenced by BMFZ

Locus: /map="9q34"

Insert length: 2380 bp  
Poly A stretch at pos. 2341, polyadenylation signal at pos. 2318

```

1  AGCCTGCGTG  CTGAGGTATG  GCGAAGCGGT  GCGGGGTCTC  TTCCGAGTC
51  TTTTCTCGGA  CGGATCTCTT  GCGGTGGGTG  TTTTCTGGGC  TGGCTGGGG
101  AGCGCGTTGT  GTTGGCAGGG  GCGCGGCGCC  GGGAGGCGCG  GGTCTCGGG
151  TGGCGCGCGG  CCGAGCGCGT  GGACGGCAGC  AGGATGGGGA  AGGCGAAGGT
201  CCGCGCTCTC  AAGCGGCGCC  CGAGCGAGCC  CGTGGTAGG  CGGGTCTGT
251  TCAGACCGCT  CAGTGGGAG  AAAAACAAGA  AGAAAAAAG  GTTTTGGAAA
301  AGCAAGCGCG  GGAAGTAG  CAGAAGCCA  GCAAGCGGCG  CCGGTCTGT
351  GGTGCGACCT  CCAAGCGCAG  CAGAAGACTT  TTCTCAAAAC  TCGAAGGCG
401  TCGAAGCTGT  CTTCTGAAGA  CAATAATCTC  AGCCCTCCGA  AAAAGCTCTT
451  GTATCTCTCT  AGATGGGTTT  CAAAAAGAAG  CCAAAATTA  TCCAGCAAAA
501  CAAAAAAGAG  ACCTCGCTCT  AAGTGAAGGG  AGAGGAGATG  CCGGACAGAA
551  AAGCCAGGA  GCGCCGCGGG  GCTCTCTCTC  CTTCAGGTTT  CAAATGAGAC
601  AGAGGCGCGC  CACTACCTCG  CACCAAGCGC  ACTGGAACAG  AGCAACAATA
701  ACATCGAGCA  TAACAGGCG  AAAGCTAAG  AGCCAGCGC  AGCCCAAGCC
751  ACCGAGGAAG  ACATCTGTT  TGACGACGTC  GACCCAGCGG  ATATGGAAGC
801  TGCCATAGGT  CCAGAGGCGG  CCAGATAGC  GAGGAACAG  TTGGTTCAGA
851  GCGAGGCGAG  COTCAGGCTC  AGCTTCTCT  AAGAGCAGG  CTTCGGCGG
901  CTGACAGAG  CATTAGCCTT  GGACTCTGAG  ATGTTGGGCG  TGGGCCCTAA
951  GGGGAGGAG  AGCATGGCGC  CCGCTGTGTC  CATCTGAAC  CAGTATGGGA
1001  AGTGCTTTA  TGACAGTAC  CTCAACCA  CTGAGCGCT  GAGGACTAT
1051  AGGACAGCG  TCACTGGGAT  TCGGCTGAG  AACTCAGC  AGGGAGAGA
1101  GCTTGAAGTT  GTTCAGAAGG  AAGTGGCAGA  GATGCTGAG  GGCAGATTTC
1151  TAGTGGGGA  CCGCTCTCAT  AATGACTTAA  AGGTACTATT  TCTTATCAT
1201  CCAAAAAGA  AGATTGGGA  CACACAGAA  TATAACCTT  TCAAGATCA
1251  AGTAAAGAT  GGAAGCGGT  CTCTGAGACT  ACTTTCAGAG  AAGATCTCTG
1301  GCGTCAAGT  CCAGAGGCG  GAGCACTGTT  CAATCAGGA  TGGCAGGGA
1351  GCAATGAGG  TGTACGTCT  GTGGAAGAG  GAGTGGGAGA  GCATGGCCCG
1401  AGACAGGCG  CCGCTGCTGA  CTGCTCAGA  CCACTGCACT  GACGACGCGT
1451  AGCATCTCT  CCGTCTCTCT  GCTGCGCC  CCGTCAAGAG  GCATGTGAC
1501  CAGTACAGG  GAGGATGAC  ATCTCCCGAG  ACTGGCACT  CTGTTAAAC
1551  CTTTTCAGAA  TCATGGCAGA  GGGCGCTGGC  GTGGTCTAC  TGAGAAGCTC
1601  CTCCTTCTCT  TTGACTTTT  GGTCTGAAC  CTGGTCTTAC  TGTCTATGTG
1651  TTTTGGGCG  CGATGTTCA  GGTGGGAG  GAGGAGCGG  CATGGGACG
1701  CCGTGGCAG  CTTTACCGAC  TGCTGACCC  CTGGGCGAG  TGAGGTTGGG
1751  GCTTGTGGG  CGGCACTCA  TACGCTGCT  TCACTGCCA  TCTTGGTGA
1801  CAGCTGGG  TGAGTCTCT  AGCACTCTC  TCTTCAGAG  CCAATTTTC
1851  CTCCTTTGT  TTAGGGACA  TAACAGCTC  TGTGGGCTT  GAGGAGCCCA
1901  GACCAAGTT  CTGAGTCTG  CTCTGAGAC  AGAGTGGCC  GGCACACAG
1951  GTTATGATC  AGGGTTTCT  TGTGGGCTT  GAACTTTCA  GCATTTATCT
2001  AATTTAAAT  GGCCAGGCT  TGCTGTGTC  GTACCCAGC  AGAGCTTCT
2051  CCGCATAGCA  CGAGGATGT  TTGCTGGGG  ACGGTGACTG  CGGTATTTC
2101  TGGAGTGGG  CAGCATGCG  AACCTTGGG  TATTTAGCT  GGAAGGCTA
2151  TGTGATGCTA  GCGGTGCTT  TCTGGGCTA  GCGCCAGTT  TGAGGCTGCC
2201  CTGGGACTA  GAGCCAGGA  CAGCCATGG  CACTGACAG  GCGACGAGT
2251  CCAAGGCTT  ATTGGGCA  CTGACCTCT  GACAGAAAG  GGGAGACAC
2301  ACCAGGATG  CGATTTAAA  TAAATGAGA  TTTTACTTG  GAAAAAATA
2351  AAAAAAATA  AAAAAAATA  AAAAAAATA

```

BLAST Results

WO 01/12659

PCT/IB00/01496

Entry NSAC2099 from database EMBL:  
\*\*\* SEQUENCING IN PROGRESS \*\*\* Genomic sequence from Human 9q34; HTGS  
phase 1, 2 unordered pieces.  
Score = 3053, E = 0.0e+00, identities = 1011/1011  
8 exons bp 104219-116190

Medline entries

95157530:  
Cloning and expression of a Xenopus gene that prevents mitotic  
catastrophe in fission yeast.

Peptide information for frame 1

ORF from 184 bp to 1449 bp; peptide length: 422  
Category: strong similarity to known protein

1 MGKARVPASK RAPSSPVAKP GPVETLTREK HXKKRPNKS KAREVSKKPA  
51 SGCAVVRPP KAPEDFSQNW KALQDWLLKQ KSQAPKPLV ISQNGSKKKP  
101 KIIOQNKRET SPQVGEENP AKQDQASRG SVPSGSKMDR KAVVPRTAS  
151 GTDHPKGTG RTNGGIVPE RGTIDNKKRK AKKAPAPPT SEDWTFQWD  
201 PADTEAAIGP EAKIARKQL QSEGSVSLS LVKEQAFGL: TRALALDCEN  
251 VGVCPKEGES HAARVALVHQ YGKCVTDIV KPTPTPTDVR TAVSGIRPEN  
301 LKQGEELVW QREYADHAG RLVGHALRW QLVVLQHP KRITADTQY  
351 KPTKSQVKSQ KPSLRLLSEK ILGLVQQAQ HCSIQDAQAA MRLTVNVKKE  
401 WESHARDRAP LLTAPDHCSQ DA

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZphtes3\_7p10, frame 1

No Alert BLASTP hits found

Peptide information for DKFZphtes3\_7p10, frame 1

Report for DKFZphtes3\_7p10.1

(LENGTH) 422  
(MW) 46671.91  
(pI) 9.79  
(HOMOLOGY) PIK:SS3818 XPM2 protein - African clawed frog 7a-96  
(FUNCTION) 03.22 cell cycle control and mitosis [S. cerevisiae, YOL080c] 2e-42  
(FUNCTION) 01.03.16 polynucleotide degradation [S. cerevisiae, YGR276c] 2e-19  
(FUNCTION) 05.04 translation (initiation, elongation and termination) [S. cerevisiae,  
YOL094c] 7a-13  
(FUNCTION) 04.05.05 mRNA processing (5'-end, 3'-end processing and mRNA degradation) [S.  
cerevisiae, YOL094c] 7a-13  
(FUNCTION) 99 unclassified proteins [S. cerevisiae, YLR107w] 6e-10  
(PROSITE) RGD 1  
(PROSITE) MYRISTYL 4  
(PROSITE) CAMP\_PHOSPHO\_SITE 2  
(PROSITE) CK2\_PHOSPHO\_SITE 6  
(PROSITE) TYR\_PHOSPHO\_SITE 2  
(PROSITE) GLYCOSAMINOGLYCAN 1  
(PROSITE) PKC\_PHOSPHO\_SITE 8  
(KW) All\_Alpha  
(KW) LOW\_COMPLEXITY 11.37 %

SEQ MGKARVPASKRAPSSPVAKPGPVETLTREKHXKKRPNKSKAREVSKKPAKSGCAVVRPP  
SEG .....XXXXXXXXXXXXXXXXXXXX  
PRD ccc

SEQ KAPEDFSQNWKALQDWLLKQKSQAPKPLVISQNGSKKKPKIIOQNKRETSPOVKGEMD  
SEG .....XXXXXXXXXXXX  
PRD ccc



[illegible]

Prosite for DKFZphtesJ\_7p10.1

PS000002	51--555	GLYCOSAMINOGLYCAN	PDOC000002
PS000003	107--111	CAMP_PHOSPHO SITE	PDOC000004
PS000004	156--160	CAMP_PHOSPHO SITE	PDOC000005
PS000005	161--165	CAMP_PHOSPHO SITE	PDOC000006
PS000006	27--30	PKC_PHOSPHO SITE	PDOC000007
PS000007	86--99	PKC_PHOSPHO SITE	PDOC000008
PS000008	96--149	PKC_PHOSPHO SITE	PDOC000009
PS000009	150--154	PKC_PHOSPHO SITE	PDOC000010
PS000010	359--362	PKC_PHOSPHO SITE	PDOC000015
PS000013	683--366	PKC_PHOSPHO SITE	PDOC000016
PS000015	367--370	PKC_PHOSPHO SITE	PDOC000017
PS000016	136--140	CK2_PHOSPHO SITE	PDOC000018
PS000017	150--154	CK2_PHOSPHO SITE	PDOC000019
PS000018	683--367	CK2_PHOSPHO SITE	PDOC000020
PS000019	368--371	CK2_PHOSPHO SITE	PDOC000021
PS000020	383--387	CK2_PHOSPHO SITE	PDOC000026
PS000026	113--147	CK2_PHOSPHO SITE	PDOC000028
PS000027	383--351	TYR_AKOSPHO SITE	PDOC000029
PS000028	113--147	TYR_AKOSPHO SITE	PDOC000030
PS000029	130--136	MYR3STYLE	PDOC000038
PS000038	151--157	MYR1STYLE	PDOC000039
PS000039	158--164	MYR2STYLE	PDOC000040
PS000040	139--145	MYR1STYLE	PDOC000041
PS000016	271--274	RGD	PDOC000016

(No Pfam data available for DKFZphtes3\_7p10.1)

DRFiphes3\_7p9  
-----

group: nucleic acid management

DRFiphes3\_7p9 encodes a novel 691 amino acid protein with similarity to human nuclear domain 10 protein NDP52.

The nuclear domain (ND)10 also described as POD or Kr bodies is involved in the development of acute promyelocytic leukemia and virus-host interactions. The NDP52 protein is part of this complex structure. In vivo, NDP52 is transcribed in all human tissues, but is redistributed upon viral infection and interferon treatment. ND10 plays an important role in the viral life cycle.

The novel protein is similar to NDP52. It contains three leucine zippers and a RGD cell attachment site. This protein seems to be a novel part of the ND819 complex.

The new protein can find application in modulation of viral infections and tumour events.

similarity to nuclear domain 10 protein NDP52

complete cDNA, complete cds, EST hits

Sequenced by BNF2

Locus: /map=329.1 cR from top of Chr12 linkage group\*

Insert length: 3003 bp

Poly A stretch at pos. 2957, no polyadenylation signal found

```
1 AAGGTGAGGG GAACAGCTGA TCCGTCTGTT GGGAGGACAG ATATCTCAAG
51 GCGAGGATGG AAGAATCACCC ACTAAGCCGG GCACCATCCC GTGCTGGAGT
101 CACCTTTCTC AATGTAGGCC GGACATACAT CCCACACACC AAGGTGGAT
151 GTCACTACAC CATTCCCCCA GGCACCATGC CCAGTGCCAG TGACTGGATT
201 GGCATCTTCA AGGTGGAGGC TCCCTGTGTT CGGATTACG ACACATTGTT
251 GTGGTCTTCC GTGGCTGAAG GTACACTGTA TGGTTCGCC ATTCACCA
301 GTTCTCAATT CCAGGCAGC TACTGCCCA AACCCAGAGC TCAGCTCTAC
351 CAGTTCGGAT ATGTGAACCG CCAGGGCCAG GTGTGTGGG AGAGCCCCCC
401 TTTCAGTTC GAGAGCCAA GGCCTATGGA TCAGCTGTA ACCTGAGG
451 AGCTCATGCG GGGCTCTGAC ATCTGTGTTG TTGTCCCAA GCGACTGTG
501 TTACAGAACCC AGCTCATGA GAGCCAGCAA GAACCGAATG ACCTGATGCA
551 GCTGAGCTA CAGCTGGAGG GACAGGTGAG AGAGCTGAGG AGCGAGTGC
601 AGAGCTTCA GAGGCTCTG GCAACTGCCA GGCAGAGGCA CAGCGAGCTG
651 ATGGAACAGT ACAAGGGGAT TTCCCGTCCC CATGGGAGA TCACAGAGA
701 GAGGAGATC CTGAGCCGGC AACAGGGAGA CCATGTGGCA CGCATCTGG
751 AGTTAGAGA TACATCTCAG ACCATCAGTG AGAAGCTGCT GACCAAGDA
801 GTGAGCTGG ACAGGCTTAG AGACACAGTG AAGGCCCTGA CTCGGGAACA
851 AGAGAGCTC CTGGGCAAC TGAAGAAGT ACAAGCAGAG AAGAGCAAA
901 GTGAGGTTA GCTTCAGTG GCACACAGG AGACACTCT CTAAATTTG
951 GACCTGAAGG AGGCGAAGG CTGGCAGAG GAGCAGAGTG CTCAGGCTCA
1001 GCGACTGAAA GACAAGGTGG CCCAGATGAA GGCACACCTA GGCAGGCCC
1051 AGCAGGGGAT GGCCTAGCTG GAGCTCTTGA AGACACAGCT TCGAGGGCC
1101 CAGGAGCTTG CAGGCTCAAG CCAGCAGAAA GCCACCTTCC TTGGGAGGA
1151 GTTGGCCAGC GCAGCAGCAG CCAGGAGCCG CACCATAGCC GAATACACC
1201 GAGCGCCGCT GGAATGGCT GAAATTAAGG CCAGGCTGGC TGAGCTTGGT
1251 TTGCACTTCA AGGAAGAAA ATCCCAATGG AGCAAGGAGC GGCAGGGCT
1301 GCTCAGAGT GTGAGGAGG AAGAGGACAA GATCTCTGAG CTGAGTGCAG
1351 AGATACCTGG ATTGACAGG GCAGTCTGAG AGGAGAGAGC CCAAGACAA
1401 GTTCTCAGA CTGAGCTGG CCGGAGAGAG GATTCTAGCC TGGTACAGTT
1451 GTCAGAAAGT AAGCGGGAGC TCACAGAGCT CGGGTACGCC CTGCTGTGCG
1501 TCCAGAGAGA AAGGAGCAG TTACAGAGAG ACBAACAGGA ATTCCTAGAG
1551 TACATGAGA AGCTTAGAGC CGGCTGGAG AAGCTGGCAG ATGAGAGTG
1601 GAATGAGGAT GCCACACAG AGATGAGGA GCGCGCTGTG GGGCTGAGCT
1651 GCGCGGAGC TGTGACAGC TCAGAGAGAG AGTCCGAGGA AGCATGAGG
1701 CTCGACCTT ATGGCTTTG TCAGGCTGGA CAGCCAGGCT CCTCTCTCC
1751 TGGGCTTCCA GAGGCTTCTC CCTTGTGTT CATCAGCAG CCGGCTCCCA
1801 TTCTCTCTCA CTTCTCTGGG CCAGCTGAGG ACAGTAGCTC TGACTCTGAG
1851 GCTGAGATG AGAATCAAT CCTCATGCA CCTCTCAAG GTGGGGGTA
1901 GGAGGCCAAC TTACTGTTT CTGAAGTGGG CAGTGCTTC TATGACATGG
1951 CCAGTGGCTT TACAGTGGT ACCTGTGTCG AAGCCAGCAG TGGGGGCCCT
2001 GCAACCCCA CATGAGAGA GTTCTTATG TGTAGGAGG CTTTCTCTC
2051 TGAGATGAC AAGGATGCC TGAGAGACA CATGATGGA CACTTCTTT
2101 TCAGCACCCA GAGCCCTTTC ACCTTTGAGT GATCTTACTC CCTGTACAT
2151 GCACAAATAC ACATCTATG ACACACAGC TCACACAGAT GCATACACTT
2201 AGGTTTCATG CCAATTTTCT ATCACACTGG GCTCATGAT ATTCTGTCC
2251 CTAAGAAGCT CTTCTGTGTT CCTGTCTTTC ATCCCAAGAT TTCTCACTTC
2301 ATCTCTCTCT ACCTGCTCTT TTTGTCCAG GAGAGGGCTC TGTGTGAGG
2351 CAGTGGCTGA ATTTATCCCC TGAAGTGGT TTTGAGGAA CCGGATGGA
2401 GGAGGCCCTC CCTGTGGGA ATAGATGCT CCACCTCAG CCTGTGTCG
```

WO 01/12659

PCT/IB00/01496

```
2451 TTCTGATACA CAGCCACTGC ACACACACAC TCACACTCAC ACTCCCTTGT
2501 CTGATGCCCC AAGGCCAHTT CCTGGGGCAC CTTACCTCTT CTTATTGGGA
2551 GTTTCGGTNG GTTACCTGCA GTTTCCTCTG GGGTCTGCAC AGAGGCAGCA
2601 GCATGGACAT CATGCCCTCT CAGGTCCCTT TTGGTTCTCA GTTTCATTGG
2651 TTCCCTCTTC TGTTCGCCCA TTGACTCTCG TGGCCACACC TACCTTTTTC
2701 CATACCTTCA GGTATTCAGT TTGGAGGGGT TTTTGTATT TTTGAGGATT
2751 CCTGTATTCT GTATCTCTCT CTGCACTCTC CTCACATGGA AAGAATAAAT
2801 GTATTCTGCG CTTCTGTGAG GAATGGGGGG AACACCTCTT CCGAGTATTC
2851 CCGATTTCGA AGGCCCCCTT CCTCTCCAG GTCCCCCACC AGCAATAAAA
2901 GCTTCCCTCT CATATCCATC CCTTTGTAGT TTGAACAAT ATATTATAT
2951 GATATTGAAA AAAAAAAA AAAAABAAAA AAAAAAAA AAAAAAAA
3001 AAA
```

#### BLAST Results

Entry HS189353 from database EMBL:  
human STS WI-11261.  
Score = 2191, E = 1.4e-92, identities = 463/485

#### Medline entries

95310349:  
Molecular characterization of NDP52, a novel protein of the  
nuclear domain 10, which is redistributed upon virus  
infection and interferon treatment.

97378672:  
Cellular localization, expression, and structure of the nuclear  
dot protein 52.

#### Peptide information for frame 3

ORF from 57 bp to 2129 bp; peptide length: 691  
Category: similarity to known protein  
Prosite motifs: RND (557-560)  
LEUCINE\_ZIPPER (163-185)  
LEUCINE\_ZIPPER (475-497)  
LEUCINE\_ZIPPER (482-504)

```
1 NEESPLSRAP SRGGVFLNV ARTIIMTKV ECHVTLPGOT MPRASDMLGI
51 FEVEAACVAD VITFMESVP ESTTGSPIN TSVOTGASTL PKAGADLIQF
101 RYVNRQGVQC GQSPFPQFKE FRPHDELVTI EADGGSDIL LVVPKATVLQ
151 MQLDESQGER NDLNQLKQL EGQVTELRSR VQLEKALAT ARGENTELME
201 QYKGISHSHE EITREKILS RQKQGVARI LELEDITQTI SRVLTREVE
251 LDRLRDTVKA LTRDZKLLG QLAEVQADKE QSEAELOVAQ QEMHMLDLQ
301 KEARSMQEEQ SAQGRLEKQK VQKMDTLQD AQORVALEVP LKQLKADQOE
351 LAASSQDQAT LQGLASASA AAKQRTIAGL KRELEVAVYQ NQRLALGLW
401 LKEKCOMSK ERAGLLQSV EAKDKILLEL AEILRLKXAV QERTOMOVF
451 RYELAREKDS SLVQLSESR ELTELASALR VLQKREQLQ EKQELGLTH
501 RELARLEKVP ADEMNQONT TEUKKAVGL SCPAALTTEE DESPEOMHLP
551 PYGLCEKGGP GSSPAGPREA SPLVVISQPA PISPHLSGPA EDSSDSSEAE
601 DEKSVIAAAV OSQGEARENL LPFLGSAFTD MABGTTVGTI SETSTGGPAT
651 PTHKEZPICK EAPFAEDNG ALZDNGOCHF FTSTQDPFTF E
```

#### BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFZptea3\_7p9, frame 3

PIR:A56733 nuclear domain 10 protein NDP52 - human, N = 2, Score = 307,  
P = 7.7e-28

TREMBL:AB008852.1 gene: "NDP"; product: "NDP52"; Bos taurus mRNA for  
NDP52, complete cds., N = 2, Score = 302, P = 4e-27

TREMBL:AC004549.1 gene: "WDGSC:H RC459W13.1"; product: "TXBP151"; Homo  
sapiens BAC clone RC459W13 from 7p15, complete sequence., N = 2, Score  
= 275, P = 2.3e-25

PIR:G02043 TXBP151 - human, M = 2, Score = 270, P = 8.5e-25

TREMBL:DW35816.4 gene: "zip"; product: "nonmuscle myosin-II heavy chain"; Drosophila melanogaster nonmuscle myosin-II heavy chain (zip) gene, complete cds., M = 1, Score = 254, P = 1.4e-17

>PIR:AS6733 nuclear domain 10 protein NDP52 - human  
Length = 446

## HSPs:

Score = 307 (46.1 bits), Expect = 7.7e-28, Sum P(2) = 7.7e-28  
Identities = 104/323 (32%), Positives = 158/323 (48%)

```
Query: 15 VNFLVARTYIPNTEVECHYTLPPGTHPSASDMIGIFVEAACVVDVHTFVSSVPESTT 74
      V F +V + YIP V CHYT +P DMIGIF+V R+Y+TF+W ++P
Sbjct: 23 VIFMSVEKFTYIPGGDVTCYHTTFTQHFIPRAKDMIGIFRVGKTKTRETITTFMVTLPIDLN 82

Query: 75 DGSPIHTSVQFOASLYPKPGALYQFRTVNRQGVCGSPFPQFREPRLDELVTLEAD 134
      + S VQF+A YLPR + YQF TV+ G V G S PFOFR D LV +
Sbjct: 83 HSAKQGVQFRTVYLPDD-ETTYQTCYVGGGVVGAASIPQFRPEHEEDILVVTTQ-- 139

Query: 135 GGSILLVVPKATVLMQ-LDES--QDENLMLQKLQLEGQVTE-LASRVQELERALA 189
      G + + K +HQ L +S Q+H HQ +LQ + + E L+S ++LE +
Sbjct: 140 GEVEKIEQHNKELCKENQELKDSCLSLQKQNSDMQAELOKQOELETQSNKKLELXVK 199

Query: 190 TAROE-HTELMQYKGISRSNGEITEERDI-LSAQGDHVARILELEDDIOTISEKVLTK 247
      + TEL+ Q K ++ E+ I + + Q + E+E +Q +E T+
Sbjct: 200 EORDYMETELL-QLKEQMKNSSEKENGIRVDQALQSLTOEKREKLVGGDQK--TE 256

Query: 248 EVE-LDLKDTYKALTREQKLLGQLEVOADKEQSAELQVQDENHMLDLKAKSW 306
      ++E L + D + EQ K +L++ +QE E QOE H DL + S
Sbjct: 257 QLEQLKENDHLFLSLTEQRKDKKLEQTVQGNKQNETTAMRKQQLNDENFDLSKLSE 316

Query: 307 QEQSAQOQLKADRVQKHTLQGAQQRV 315
      E OR R+++ D L + R+
Sbjct: 317 NEIICHALQQRERLEGENDLERENSRL 345
```

Score = 304 (45.6 bits), Expect = 2.1e-27, Sum P(2) = 2.1e-27  
Identities = 98/337 (29%), Positives = 163/337 (48%)

```
Query: 15 VNFLVARTYIPNTEVECHYTLPPGTHPSASDMIGIFVEAACVVDVHTFVSSVPESTT 74
      V F +V + YIP V CHYT +P DMIGIF+V R+Y+TF+W ++P
Sbjct: 23 VIFMSVEKFTYIPGGDVTCYHTTFTQHFIPRAKDMIGIFRVGKTKTRETITTFMVTLPIDLN 82

Query: 75 DGSPIHTSVQFOASLYPKPGALYQFRTVNRQGVCGSPFPQFREPRLDELVTLEAD 134
      + S VQF+A YLPR + YQF TV+ G V G S PFOFR +E
Sbjct: 83 HSAKQGVQFRTVYLPDD-ETTYQTCYVGGGVVGAASIPQFR--PEDE----- 130

Query: 135 GGSILLVVPKATVLMQLODESQDERMLQKLQLEGQVTELASRVQELERALATAROE 194
      DILV Q +++E +Q +L + +L+ L+ + ++ L +QE
Sbjct: 131 --EDILVVT-----QGEVEEIEQHNKELCKENQELKDSCLSLQKQNSDMQAELOK-QOE 182

Query: 195 HTELMQYKGISRSNGEITEERDI-LSAQGDH-VARILEDDIOTISEKVLTKVELDR 253
      E ++ I ++ ++ ++Q D+ +L L++ Q +S + + +D+
Sbjct: 183 ELLETQS-----INRKLKLVKQDWTETELQLKQKNSSEKENGIRVDQ 232

Query: 254 LRDYKALTREQKLL--QLKEVQAD---KEQSAELQVQDENHMLDLKAKSMOE 308
      L+ + +E EKL- Q R Q + KE L + +Q L+ + + Q
Sbjct: 233 LQALSTQEKREKLVGGDQDTQLEQLKENDHLFLSLTEQRKDKKLEQTVQGNKQ 292

Query: 309 EQSA--QADLKADRVQKHTLQGAQQRVAALEPLFLQALQAE 351
      E A + Q L D+ + L + + L+ KE+L G +L
Sbjct: 293 EYTAHKKQQLNDENFDLSKLSENEIICHALQQRERLEGENDL 337
```

Score = 124 (18.6 bits), Expect = 2.3e-06, Sum P(2) = 2.3e-06  
Identities = 53/227 (23%), Positives = 113/227 (49%)

```
Query: 138 DILLVVPKATVLMQLODESQDERMLQKLQLEGQVTELASRVQELERALATARDITE 197
      DILV Q +++E +Q +L + +L+ L+ + ++ L +QE E
Sbjct: 132 DILVVT-----QGEVEEIEQHNKELCKENQELKDSCLSLQKQNSDMQAELOK-QOELE 185

Query: 198 LMEQYKISRSNGEITEERDI-LSAQGDH-VARILEDDIOTISEKVLTKVELDR 256
      ++ I ++ ++ ++Q D+ +L L++ Q +S + + +D L+
Sbjct: 186 TLQS-----INRKLKLVKQDWTETELQLKQKNSSEKENGIRVDQ 235

Query: 257 TVKALTREQKLLGQLEVOADKEQSAELQVQDENHMLDLKAKSMOEQSAQOR 316
      + E EKL- VQ D++++E +L+ ++EN HL L L E + O++ ++
Sbjct: 236 QLSTQEKREKLVGGDQDTE-QLEQLKENDHLFLSLTEQRKDKKLEQTVQGNKQ 288
```



[FUNCAT] 01.05.04 regulation of carbohydrate utilization [S. cerevisiae, YBR289w]  
4e-06  
[FUNCAT] 04.05.01.04 transcriptional control [S. cerevisiae, YBR289w] 4e-06  
[FUNCAT] 03.19 recombination and dna repair [S. cerevisiae, YML250w] 4e-06  
[FUNCAT] 03.13 meiosis [S. cerevisiae, YML250w] 4e-06  
[FUNCAT] 1 genome replication, transcription, recombination and repair [M. jannaschii, MJ1643] 1e-05  
[FUNCAT] 98 classification not yet clear-cut [S. cerevisiae, YJR134c] 4e-05  
[FUNCAT] 11.04 dna repair (direct repair, base excision repair and nucleotide excision repair) [S. cerevisiae, YKR095w] 4e-05  
[FUNCAT] 08.19 cellular import [S. cerevisiae, YML243w] 7e-05  
[FUNCAT] 01.02.16 polynucleotide degradation [S. cerevisiae, YML243w] 7e-05  
[FUNCAT] 06.10 assembly of protein complexes [S. cerevisiae, YML243w] 7e-05  
[FUNCAT] 08.99 other intracellular-transport activities [S. cerevisiae, YML079c] 2e-04  
[FUNCAT] 03.01 cell growth [S. cerevisiae, YML079c] 2e-04  
[BLOCKS] BL006828 1P domain proteins  
[ICI] 3.4.1.32 myosin ATPase 1e-13  
[PIRKEW] nucleus 6e-10  
[PIRKEW] phosphotransferase 2e-07  
[PIRKEW] duplication 3e-07  
[PIRKEW] citrulline 1e-09  
[PIRKEW] tandem repeat 1e-13  
[PIRKEW] heart 5e-11  
[PIRKEW] endocytosis 5e-09  
[PIRKEW] polymorphisms 3e-06  
[PIRKEW] cornified cell envelope 1e-06  
[PIRKEW] transmembrane protein 6e-12  
[PIRKEW] serine/threonine-specific protein kinase 2e-07  
[PIRKEW] cell wall 1e-06  
[PIRKEW] zinc finger 3e-09  
[PIRKEW] metal binding 3e-09  
[PIRKEW] dna binding 1e-08  
[PIRKEW] muscle contraction 1e-11  
[PIRKEW] igg constant region-binding 1e-06  
[PIRKEW] acetylated amino and 4e-09  
[PIRKEW] actin binding 1e-13  
[PIRKEW] mitosis 9e-05  
[PIRKEW] microtubule binding 9e-09  
[PIRKEW] ATP 1e-13  
[PIRKEW] thick filament 1e-10  
[PIRKEW] phosphoprotein 1e-13  
[PIRKEW] epidermis 1e-06  
[PIRKEW] leucine zipper 1e-07  
[PIRKEW] glycoprotein 4e-07  
[PIRKEW] skeletal muscle 4e-10  
[PIRKEW] disulfide bond 1e-07  
[PIRKEW] calcium binding 1e-09  
[PIRKEW] alternative splicing 1e-10  
[PIRKEW] coiled coil 1e-13  
[PIRKEW] P-loop 1e-13  
[PIRKEW] heptad repeat 6e-10  
[PIRKEW] methylated amino acid 1e-13  
[PIRKEW] basement membrane 3e-06  
[PIRKEW] immunoglobulin receptor 2e-07  
[PIRKEW] peripheral membrane protein 5e-09  
[PIRKEW] dimer 1e-07  
[PIRKEW] cardiac muscle 1e-10  
[PIRKEW] extracellular matrix 3e-06  
[PIRKEW] hydrolase 1e-13  
[PIRKEW] microtubule 6e-10  
[PIRKEW] muscle 2e-09  
[PIRKEW] membrane protein 3e-06  
[PIRKEW] EF hand 1e-09  
[PIRKEW] cytoskeleton 6e-12  
[PIRKEW] hair 1e-09  
[PIRKEW] calmodulin binding 5e-09  
[PIRKEW] Golgi apparatus 3e-09  
[SUPTAM] myosin heavy chain 1e-13  
[SUPTAM] conserved hypothetical P115 protein 1e-08  
[SUPTAM] hypothetical protein YK1074c 5e-07  
[SUPTAM] centromere protein E 9e-09  
[SUPTAM] unassigned Ser/Thr or Tyr-specific protein kinases 2e-07  
[SUPTAM] calmodulin repeat homology 1e-09  
[SUPTAM] myosin motor domain homology 1e-13  
[SUPTAM] alpha-actinin actin-binding domain homology 3e-13  
[SUPTAM] tropomyosin 3e-07  
[SUPTAM] plectin 3e-13  
[SUPTAM] trichomyosin 1e-09  
[SUPTAM] plectatrin repeat homology 4e-06  
[SUPTAM] ribosomal protein S10 homology 3e-13

959

WO 01/12659

PCT/IB00/01496

Prosite for DKFZphtes3\_Tp9.3

PS00005	190->193	PKC_PHOSPHO_SITE	PDCC00005
PS00005	241->244	PKC_PHOSPHO_SITE	PDCC00005
PS00005	257->260	PKC_PHOSPHO_SITE	PDCC00005
PS00005	468->471	PKC_PHOSPHO_SITE	PDCC00005
PS00005	652->655	PKC_PHOSPHO_SITE	PDCC00005
PS00005	667->670	PKC_PHOSPHO_SITE	PDCC00005
PS00006	28->32	CK2_PHOSPHO_SITE	PDCC00006
PS00006	43->47	CK2_PHOSPHO_SITE	PDCC00006
PS00006	68->72	CK2_PHOSPHO_SITE	PDCC00006
PS00006	72->76	CK2_PHOSPHO_SITE	PDCC00006
PS00006	129->133	CK2_PHOSPHO_SITE	PDCC00006
PS00006	156->160	CK2_PHOSPHO_SITE	PDCC00006
PS00006	208->212	CK2_PHOSPHO_SITE	PDCC00006
PS00006	239->243	CK2_PHOSPHO_SITE	PDCC00006
PS00006	282->286	CK2_PHOSPHO_SITE	PDCC00006
PS00006	305->309	CK2_PHOSPHO_SITE	PDCC00006
PS00006	376->380	CK2_PHOSPHO_SITE	PDCC00006
PS00006	383->387	CK2_PHOSPHO_SITE	PDCC00006
PS00006	468->472	CK2_PHOSPHO_SITE	PDCC00006
PS00006	520->524	CK2_PHOSPHO_SITE	PDCC00006
PS00006	537->541	CK2_PHOSPHO_SITE	PDCC00006
PS00006	539->543	CK2_PHOSPHO_SITE	PDCC00006
PS00006	543->547	CK2_PHOSPHO_SITE	PDCC00006
PS00006	593->597	CK2_PHOSPHO_SITE	PDCC00006
PS00006	595->599	CK2_PHOSPHO_SITE	PDCC00006
PS00006	597->601	CK2_PHOSPHO_SITE	PDCC00006
PS00006	612->616	CK2_PHOSPHO_SITE	PDCC00006
PS00006	639->643	CK2_PHOSPHO_SITE	PDCC00006
PS00006	652->656	CK2_PHOSPHO_SITE	PDCC00006
PS00006	667->671	CK2_PHOSPHO_SITE	PDCC00006
PS00006	683->687	CK2_PHOSPHO_SITE	PDCC00006
PS00008	39->45	MYRISTYL	PDCC00008
PS00008	107->113	MYRISTYL	PDCC00008
PS00008	204->210	MYRISTYL	PDCC00008
PS00008	414->420	MYRISTYL	PDCC00008
PS00008	561->567	MYRISTYL	PDCC00008
PS00008	613->619	MYRISTYL	PDCC00008
PS00016	557->560	RGD	PDCC00016
PS00029	163->185	LEUCINE_ZIPPER	PDCC00029
PS00029	475->497	LEUCINE_ZIPPER	PDCC00029
PS00029	482->504	LEUCINE_ZIPPER	PDCC00029

(No Pfam data available for DKFZphtes3\_Tp9.3)



DKF2p154e3.8e24  
-----

group: signal transduction

DKF2p154e3.8e24.3 encodes a novel 658 amino acid putative GTP-binding protein, related to yeast YGL099w and mouse H9RL putative GTP-binding proteins.

GTP-binding proteins are involved in various signal transduction pathways, transferring the signal of a cellular receptor to an intracellular signal cascade.

The new protein can find clinical application in modulating/blocking the response to a cellular receptor.

strong similarity to guanine nucleotide binding proteins

complete cDNA, complete cds, potential start at bp 31, EST hits

Sequenced by MediGenomix

Locus: unknown

Insert length: 3290 bp

Poly A stretch at pos. 3249, polyadenylation signal at pos. 3251

```
1  COTCAGCGG TCGTGTGCG ATGGGCGGA GGAGAGCCCC GGGCGTGGG
51  TCGCTGGAG GGGGCTTAT GCCCATCAG ACTCAGCGA GCGGAGCGA
101 TCTTCACAT GACTCTGCT TGCACACAG TGAATCTAT GATGGCTTG
151 ATTGGGGTC TCTTAATCT CAGTCAGTA CTGAACAGG CTCCTTGAT
201 GACTTGTCT TACTGACAG ACTTCAGAG ACAGGTTTG TACTGAAA
251 ACTTAATAT AGCTTTGCG CTGCTAGGC TAGACATGA CTACTGTCT
301 TCGAGGAGG CCGAGAATT AGAAGCTCC ATGAAGAAA CAACAGCTC
351 TTGTGTATG CGAGGAGGC AATCGGAG CAATATATA CCGACAGA
401 ACTCAACAA CGAGAGAAG ATAATCTCT AGAATGGAG COTCAGCTG
451 TCGCGCTAG AGAGGACAG AGCTGATAT TGACTCCAT TGAACDAA
501 TCGACTTTT GGGCGAGCT CTGGAGACT ATTGAGAGA GTATATTCT
551 GGTCCAGAT GTAGATGCT GAACCCCAT CCTGTTTGA TGTAGGATT
601 TGAATGTTA TGTGAADA ATGATGCGA ATAGGAGA COTATTCTG
651 ATCAACAGC CAGACTGCT GACTGCTAG CAGCGAGTC CCTGGGCAT
701 GTACTTCGA AAGAAGATG TCAAGCTAT TTCTGTGTA GCTTGGCGC
751 GAGCATTCG CTTGAATGT GACTTCGAG AGAGGCGAA CAGATGAT
801 AGACAAGCA ACACAATGA GTTTGACAT TCCAGTTTC ACCAGCTGA
851 AATTGCCAC AGTCATCCG AACATCTCC AGTAGGGAT TCTCTTCAC
901 TTAGTAAAA TCCACAGC GATGAATAT ACATGATTT TCGAGCTCT
951 CCAGAGGAG AGGAAGACA CTGGCAGAG TGCTCAGAG AAGACGTC
1001 CAGGAAGAG GACTCAGCT AGGACTGGA GAAGACTCT ATCGAGATT
1051 CTGAGCTCG GAGGAGAAA ACCCGACGA AGAGGKCAT ACAGATTTT
1101 AGCCATCTG TATCCAGCA GGAGTTACT GAGCTCTTT AGGAGCTCA
1151 CACTGGAGA AGGTGAAG ATGGCAACT TACGCTGGA CTGTGGGCT
1201 ACCCTAATG TGGTAAGAT TCAACATCA ACACCATAT GGGCAGCAG
1251 AAGATATCT TGTCTGCC ACCTGCTAC ACAAGCACT TTCAGACTCT
1301 CTATGGGAG CTGGGCTCT GCCTGTGGA CTCTCTCTG TTGCTATGC
1351 CATCTTTCT GTCTACCAAG GCAGAAATGA CTGACAGCG AATCTCCCA
1401 ATTGATCAG TGAGATCTA TTTCTCTCT GTATCACTAG TTGTCAGAA
1451 TATTCCAGA CATCTTTAG AGCTACTTA TGGCATTAG ATCAATGCG
1501 CTAGAGAGA TGAAGATCC CACCGACTC CAACATCGA AGAATCTTG
1551 ACAGCTTATG GATACATCG AGATCTATG ACAGGCTG CACAGCGCA
1601 CAGCTCTGA TCTGGGCTC ACATCTTGA GGACTATGT AGTGTAMGC
1651 TGTCTACTG CATCTCTCT CTTGGAGAG ATCTGTATC TTTTCAGAT
1701 CACACACAG GACTCTAGA GAGCAATAT ACAGTATG AATTAATAT
1751 GCAGCTAGC AGAATATAA AAGCAAGCA GATTGAAAT ATCTTGACA
1801 AAGCTTTTT CATCAAGAG AATGTAGGG CTTTGACCA AGAGTTCAG
1851 CTTGTATGG GTTACAGCT CGGAGTGTG GTACTGACT CATCTCTGC
1901 GAGCTCTGAG AACGGGGCG GGAAGGCTG GAAAAACAT GGCACAGAA
1951 ATMAAAGAA AAAAGTGTG AGACTCTGA AGACCTGGA TATGTAGGT
2001 TGGGTCCAA CAGAAATGT ATTCTGATG TGCAGATGA AAGAGAGGA
2051 AGTGTGCTT TGGCTGTGA ACTGTCCAA GACACTAGA CTTAGAACG
2101 GGCTCTGCT TTGAGAGA CGGCTGACC CACAGCTCT CATGTAGAA
2151 CCAAGGGCT CTTGGAACA CCGCTCTGA CAAAAGGAG TCATCTGGA
2201 GCGCGAAGT CTAATCTCT GCGCGGACA GTGGCTCAG CACCAACATG
2251 GAGAAAGCG GTCTCTACT AATATACAA AATATTAGT AGGCTGCTG
2301 GCGCGAGCT GTAATCTCG CTACTCTGA GCGTGAGCA GGAATATCA
2351 TTGAACGAG GAGGAGAGT TTGAGTGA TGGAGATGC GCGCTGCMC
2401 TCAAGCTCG GGCACAGCT GAGACTGAT CACAGAAAA AATTTTGA
2451 AGGCTGCTT CACGAGAC ATTTGGGAG AGGTGAAG AGAATTCOC
2501 CATCTGAGT GTCTGATTC GTTCTCTCC ACTTACAA AGGACTTGC
2551 CTACCTTAG TTACAGCG GGGCTACCT CCACTAAGA ACATGTAGA
2601 TGTACACAC GGTGAGAGG GAAGCTGCG TATTACTAC CTAGCCCCA
```

WO 01/12659

PCT/IB00/01496

```
2651 TCTTCACCTGG TTATTCACCT TATTAAAT GTCCAGATA AGCAATCTC
2701 CATATAGAGG AAGTAGATTA GTGTTGCTT CGGATGGGA GGAATGGGA
2751 GATTGAGTC TTCTTTTGG AGTATAAA ATGCTCTAA ATTGACTTA
2801 GCGATGCTCA CACAACCTG AATATGCTA AGACCATTA ATTACACCT
2851 TTACGTTGCT GAATTGATG GTATGTAAAT TATAGTTCA TACATAGTT
2901 ACAAAGATA ATCAAAGGA TGAAGGACT ATTGATGAG TTGATCTG
2951 TGTCTCACC GAGTCTCATG TTGAAATGA AGCCCTCTG TGGAGGCGA
3001 TGGGATTATG GGGCAGAGT CTCACAAAGC GTTAGCACC ACCGCTCAG
3051 TGTCTCTCTG CTGATATCT CTTCTCTCTA CATCTCTCTG CTGAAATG
3101 TGTGTCCTG CCCTCTCTG TCCCTCTCTG TCTGCGCATA TAAGATCTG
3151 CTGCTCTCTG TTGCGCTCTG AACTGATGTC TACGTTCTCT GAGGCTCTG
3201 TACAGGAAA AGCTCTCTG CTCTCTCTG CATCTCTG ACCGTGAGCC
3251 AATTAAACCT CTTTCTTTA TAAAAAAA AAAAAAAGG
```

#### BLAST Results

No BLAST result

#### Medline entries

No Medline entry

#### Peptide information for frame 3

ORF from 21 bp to 1994 bp: peptide length: 658  
Category: strong similarity to known protein

```
1 MGRARAPAGG SLGRALKOHO TORSRSHKHT DSWLMTSELM DGYDWGRML
51 QSVTEQSSLD DFLATLALAG TEFVAERLM KPVPALARTG ILSFRESQRI
101 ELHLEQTFQ LCIPSRFHH QATTREESLD AKIOWFLDM RLVMLLEED
151 KLILTFPERM LQPRQLMHV IERSDIVVQI VDAHNPLLP CEDLECTVRE
201 HDWKERVIL INKADLTAE QRSAMHPTF KEDVIVPMS ALAGALPLNG
251 GSELEANRDD QSNHTTEFGH SSFOQALSH SESELPARO SPSEINPTT
301 QEDDSYENG PREZEDDMOT CREEDCFREE QCSQWRESS TAGSEARSRK
351 TPQKQJLHH SILVSPQELL ELTFELFGR SVEDCOLTMO LVGVNPKES
401 STINTINGMK KVSVSATPGH TKHPOTLYVE PGLCLDCPG LHPSPVSTR
451 ADHTCSGILP IDKRDHVPF VSLVQCMFM HVLEATYGIH ITPREDEDP
501 HSPFTSEILL TAYOTMCPH TANGUDOPM BAVILEQDY SELLICHPF
551 PGROPPTFGH QHRLLEHMK WSOEIQMLG RKKAKOIEH IVDRTFFHGE
601 HVKALTKGVQ AVHGTPOSG VVTASTASSE HGAEPWKKH GRHRRKESR
651 RUTELDM
```

#### BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DKFiptes3\_Be24, frame 3

SWISSPROT:YANG\_SCHPO HYPOTHETICAL GTP-BINDING PROTEIN C3F10.16C IN CHROMOSOME 1., N = 3, Score = 560, P = 1.6e-111

PIR:S64106 hypothetical protein YGL099w - yeast (Saccharomyces cerevisiae), N = 2, Score = 344, P = 2.6e-105

TRINHL:CEAF3143.1 gene: "C53H9.2"; Caenorhabditis elegans cosmid C53H9., N = 1, Score = 551, P = 2.9e-53

SWISSPROT:NMRL\_MOUSE POSSIBLE GTP-BINDING PROTEIN NMRL., N = 2, Score = 311, P = 7.5e-31

>SWISSPROT:YANG\_SCHPO HYPOTHETICAL GTP-BINDING PROTEIN C3F10.16C IN CHROMOSOME 1.  
Length = 616

#### RSPs:

Score = 560 (84.0 bits), Expect = 1.6e-111, Sum P(3) = 1.6e-111  
Identities = 119/253 (47%), Positives = 162/253 (64%)

Query: 12 LGRAKMKTORSRSHKHTDSWLMTSELMQGYDWGRMLQSVTEQSSLDLFLATLALAGT 71  
LGRA+ T+ R+ + N+ + R L+SVT ++ LD+FL TAEI

Seqct: 12 LGRAIOSDTKNNK--GGLKHIWDSOFK--KAALSVYTHETDDEFMTAEIGEV 67  
Query: 72 EFVAEKLIKFPV--AEARTGLSEKESQRIKLEHCHOFICIPRPFVHMQHTPEELAQ 130  
EF+AEK M+ + E LLS EE+ R K+ E+HK L IPRRP+H+Q TT EL +  
Seqct: 69 EFIAEKHVTIOMFEQHPFLSEKAAKSKQKQKHKRLTIIPRPHMQHTTAVELOR 127  
Query: 131 AEKDFLEVRROLVLEEEQRLITFFERKLPWQLWVIERSDIVVQIVDARUPLFR 190  
E++FL NRR L -L++ + I-TFFERKL+ WQLWVIERSD+VQIVDARUPL FR  
Seqct: 128 HEKESFLERKULALQDQVCEIYTPFERLEHLEHQLWVIERSDIVVQIVDARUPLFR 187  
Query: 191 CELECYVEDGDANEMVILLHKAADLTADQRSANMYFEREDKVTIWSALAGIPLWG 250  
LE VYEC+ +R+H +L+HEAD-LT EQR+ M+ YF + + +PISA A H  
Seqct: 188 SAHLEQYVEKVGPSKFNLLVHKAADLTEDQRYHSSYFVNDHIFPLFFRANMAA-EANE 246  
Query: 251 DSEKKAHRRDQSH 264  
E+ + SM  
Seqct: 247 RGEDETVESTSSM 260  
Score = 532 (79.8 bits), Expect = 1.6e-111, Sum P(3) = 1.6e-111  
Identical = 131/323 (40%), Positives = 192/323 (59%)  
Query: 340 STADSEARSRTKPNQIWHFSLVSKQELLEFEKLTGRVKGQ--LTVGLGVYHW 397  
ST+ +E + +H+ S + + + L +F+ + + + DG+ +T GLVGVYHW  
Seqct: 256 STSNEI PESLOADENVHS-SRIATLVLEGIFEFAS--TLPOGRTMTGLGVYHW 312  
Query: 398 GKSSTINTINGKKSVSATPQTKHFTLYVEPGCLCDPGLVHPSFVSTKADHTCSG 457  
GKSSTIN +G+KKSVS+TNG TKHFT+ + + L DCPGLV PSF +T+H+ + G  
Seqct: 313 GKSSTINALGKKSVSATPQTKHFTLYVEPGCLCDPGLVHPSFVSTKADHTCSG 372  
Query: 458 ILPIQKHQVPPFSLVQWIPWVULATYGIH-ITPREDPMPPTSEELLTATYGN 516  
+LIDH+ + P +L+ + IP+ VLE Y I I I P E E P+ +E+L +  
Seqct: 373 VLPIDQLEYTPSALNAERIPEVLETLTIRIRIRPIE-EGDTGVPSAGEVLFPFAS 431  
Query: 517 RGFMTAH-GPDQPSRARTILEDVSGKLYCHPFPQ--RDPYTFQHQHQLLEHRRMSD 573  
RGFM AH G PD R+AR -LKDVT+KLLY HPFF F +H + + + SD  
Seqct: 432 RGTWRAHGTGDSAAARILLQDYGKLLVHPPFVHSGSEFEDHDKLVSA-TSD 490  
Query: 574 EIKHQLGR--HKKAKQIDH-IVKTFYHGH--VRALTEQVAVH-G-TEPSSGVVTA 624  
I +L R + + E+ +VD +F CDH VH + RG G YK + +  
Seqct: 491 BITEKLTORTAISDHTLSAESQVDDQTF-CEPHVAVHVGTAHAGPVTYGRNTHPFF 549  
Query: 625 STASSENAGK-PWKNKRNKKEKSRRL 652  
+++ + K P G + E+H+L  
Seqct: 550 ORALMODASPKYPMAGGRPLSRRAKRL 578  
Score = 47 (7.1 bits), Expect = 1.3e-60, Sum P(3) = 1.3e-60  
Identical = 21/84 (25%), Positives = 33/84 (41%)  
Query: 552 GRPVTYQHQRLLKNNSEIKHQLGKAKQIDH-IVKTFYHGHVRALTEQVGA 611  
G D T++ + +DE + R K +E I +R F TK  
Seqct: 248 GEDLETYSTSMEI PESLOADENVHSSRIATLVLAGIYER--FASTLPDQRTKHTFG 305  
Query: 612 VNGYKPGSGVVYTAASSENAGK 635  
+GY P G +ST ++ G+ E  
Seqct: 306 LWGY-PMYG--KSTIALVGSSE 326  
Score = 43 (6.5 bits), Expect = 1.6e-111, Sum P(3) = 1.6e-111  
Identical = 7/13 (53%), Positives = 9/13 (69%)  
Query: 638 KKHGKRNKKESS 650  
KKH +HK+ K R  
Seqct: 596 KKHKKRNKKEQR 608

Pendant information for DRFphtes3\_8e24, frame 3  
-----

Report for DRFphtes3\_8e24.3

[LENGTH] 658  
[MW] 75226.58  
[pI] 5.86  
[MOLEC] SWISSPROT:YANG\_SCHPO HYPOTHETICAL GTP-BINDING PROTEIN C3F10.14C IN CHROMOSOME  
I. 5e-56  
[FUNGAT] 99 unclassified proteins [S. cerevisiae, YC1099w] 3e-55  
[FUNGAT] 1 general function prediction [M. lamarischi, M21464] 1e-16  
[FUNGAT] 08.16 extracellular transport [S. cerevisiae, YEN006w] 3e-09  
[PIKIN] P-loop 1e-27  
[FININ] GTP binding 1e-27  
[SUPFAM] conserved hypothetical protein MG442 7e-08

```

Protein for OFFFphenL_Sec2_1
P000001 264-268 ASN GLYCOSTYLATION P00C00001
P000001 359-363 ASN GLYCOSTYLATION P00C00001
P000001 410-414 ASN GLYCOSTYLATION P00C00001
P000003 21-232 PGC_PHSF30_SITE P00C00003
P000003 21-232 PGC_PHSF30_SITE P00C00003
P000003 21-232 PGC_PHSF30_SITE P00C00003
P000003 97-100 PGC_PHSF30_SITE P00C00003
P000003 348-351 PGC_PHSF30_SITE P00C00003
P000003 378-381 PGC_PHSF30_SITE P00C00003
P000003 448-451 PGC_PHSF30_SITE P00C00003
P000003 493-496 PGC_PHSF30_SITE P00C00003
P000003 514-516 PGC_PHSF30_SITE P00C00003
P000003 541-544 PGC_PHSF30_SITE P00C00003
P000003 639-652 PGC_PHSF30_SITE P00C00003
P000003 654-656 PGC_PHSF30_SITE P00C00003
P000006 57-61 C12_PHSF30_SITE P00C00006
P000006 93-97 C12_PHSF30_SITE P00C00006
P000006 113-117 C12_PHSF30_SITE P00C00006
P000006 153-159 C12_PHSF30_SITE P00C00006
P000006 252-256 C12_PHSF30_SITE P00C00006
P000006 279-283 C12_PHSF30_SITE P00C00006
P000006 325-329 C12_PHSF30_SITE P00C00006
P000006 352-356 C12_PHSF30_SITE P00C00006

```

WQ 01/12659

PCT/IBRU/01496

PS00006	281->285	CK2_PHOSPHO_SITE	P00C00006
PS00006	293->297	CK2_PHOSPHO_SITE	P00C00006
PS00006	299->303	CK2_PHOSPHO_SITE	P00C00006
PS00006	305->309	CK2_PHOSPHO_SITE	P00C00006
PS00006	320->324	CK2_PHOSPHO_SITE	P00C00006
PS00006	322->326	CK2_PHOSPHO_SITE	P00C00006
PS00006	340->344	CK2_PHOSPHO_SITE	P00C00006
PS00006	365->369	CK2_PHOSPHO_SITE	P00C00006
PS00006	449->453	CK2_PHOSPHO_SITE	P00C00006
PS00006	493->497	CK2_PHOSPHO_SITE	P00C00006
PS00006	505->509	CK2_PHOSPHO_SITE	P00C00006
PS00007	480->488	TYR_PHOSPHO_SITE	P00C00007
PS00007	190->198	TYR_PHOSPHO_SITE	P00C00007
PS00008	9->15	MYR1STYL	P00C00008
PS00008	432->438	MYR1STYL	P00C00008
PS00008	620->626	MYR1STYL	P00C00008
PS00009	1->5	ANIGATION	P00C00009
PS00009	378->382	ANIGATION	P00C00009
PS00017	393->401	ATP_GTP_A	P00C00017

(No Pfam data available for DRF2phtes3\_Ba24.3)

## DRFphea3\_8q11

group: testes derived

DRFphea3\_8q11 encodes a novel proline-rich 919 amino acid protein without similarity to known proteins.

The novel protein contains an ATP/GTP-binding site motif A (P-loop).  
No informative BLAST results; No predictive prosites, piam or-SCOP motifs.

The new protein can find application in studying the expression profile of testis-specific genes.

unknown, prolin rich protein

1 EST hit (from testis library)

Sequenced by MediGenomix

Locus: unknown

Insert length: 3100 bp  
Poly A stretch at pos. 3036, polyadenylation signal at pos. 3041

```
1 AGAGTCTTCC CTCAGCATAT TTTACGATAG AGAGATCTTT GTTCCAAATGG
51 AAGAAGGTGA GGAATCAGAG AGTGAATCCC AGACAGGAT TTCTGAGTCC
101 CACACTCCCC TCAGCCAAAA TTATCTTTCC GAGGCCAAGA CTGACTCTTC
151 AGACAGCTTC CAGTTCTGAG AGATCTCGCA GCTAAAATA GCGCAAAAC
201 TCTTAAGGAG TCAATATACC CCGGATGTCG CTCACCTCTT AGCTTCAGGT
251 CTGCTCTTAA AATAGCTTAT CTGGCTAGAG TTGTGCCAT GTTCAGAGCT
301 TAATTCCTAT CATAAATTAC AGACCACTTC GGCGCCTTAT CTCTTTATCT
351 ATCLACAGCT CGACTCTGTA GGCATCTCTG AAGGECATGG TGAGGTTCGG
401 TTGATCTTTC GCTTTAGCTT GAGATTTGGG AAAGATCCCC AATCTTCAAA
451 GTATCGTGAA AGAGATAGAG CCGTCATAGC GAGAGGCCCT ATATACCAT
501 CACAGAGGAA AGCTAAATAT TATAGTCAAG CTTCAGAGAG TCCATCTTCC
551 ACAATGATT TGCATCTGG GCTTCCCGAG TCCCTGCTC CTGTACAAGT
601 CTACATCAGG CGAGGACAGC GCAGCAGGCC TGACTTAGTA GAAGAAGCAA
651 AACTTAGAGC ACTGGGCGC TATCAATTCA CTCAGCTTCA CAATCTTCCA
701 GAGAGTCACT CTGAAGGAGC TCAGANTDAA AAACGGCTTA AGCTGAGAAC
751 CAAGAAGACC TGTATTATCA AATATGCAAT GAAGAGATCT ACCAGCGCAC
801 TTGAAGAAC CAAGAGCTTC TACACAAACA GTAGAACAG AATAGAGAGT
851 CTTCTAGGG AATTAGAGC CCATTTAAGA AGGAGAGGGA TTGGAGCAAC
901 TCACAGAGAT AGTGCCTTTT TAAAGAGACA ACTTAGAATA CTTTCCACAC
951 CCAAGTTTAT GCACTGCTTT TTTAGAGGCC TAAAGCGGGC ATTCCAAACA
1001 GCACACAGAG TTATAGCTTC TTTHGGGGGG AAGCTTGTGG AGCGAGCAGG
1051 GCGACACANT TTCTGGGCAA GCAAAATCTA TTATCCAAA CAAGATCCCA
1101 GGAATATTTC TTACCCAGC AGTATCAAAA GAGACAGAG CTACAGCTGAC
1151 AAGCTAGGCG CAGCAGGCTC AACCAATTAG CAGAGAGACA TATTTGCGG
1201 AGAAGGGTC CAGTGCAGAT CAGCTCAACA GCGAAGAGA GCTTACTCTT
1251 TCCACCCAG ACTCTTTGCA TTGCCCAAGC CCACAGATTC CEAAGTGGT
1301 ATGCTTTTCC AAATGCTCTC AGTGGGGCAG CCTCTAGAAA CTCTTCAAA
1351 GGCAGTNOT AGCAGATCAA AGAAAACCTT CTATAGAAT GAACCTTCCA
1401 GCGAGGATCT TAGAGACTTC TCACACAGC GAAACAGAGT TCAGGTCGGA
1451 GGAAGATCC TACTGGTTC CCGTGTGAG AGACCTGGG ACCGACATCT
1501 TAAAGACAAA CTCACACACA AGGAGCATAA CCACCCAGC TTCTATAGGG
1551 AGCAACCCC AGCCGCTCTT TCTGAGAGA CCCCTATATA CCGCTCTGG
1601 AGAACCATC GCACTCCCTC TGAGAGAGAG CAGCCAGTTT CTTTGGAGAG
1651 AGACATCAG ACTCCCTCTC AGAGAGACCA CTGCATCCC TGTAGAAA
1701 ACATCTCAG TCTCTCTGAG AGAAGCTGGC GCAGTCCGTC TCAGAGAAAT
1751 CACTGCACTC CCCCCAGAG GAGCTGTGAC AGTCTCTCTG AAGGGGGGCT
1801 TCACAGTCCC TCACAGAGA CCAATGCGGG TCGCTCTGAG AGACAGATC
1851 ACAGTCCCTC AGAGAGAGC CATGCACTC CCTCAGAGAG AAGCCATCCG
1901 AGTCCCTCTG AGAGAGACA TCGAGTCCC TCGCAGAGA GCACTCTCGG
1951 TCCCTCAGAG AGAGGCATT GCAATGCTTC TGAGAGAGA CATGCGATC
2001 CTTCTCAGAG GAGCCATGCT GTTCCCTCTG AGAGAGACA TCACATCCC
2051 TTTAGAGAA GCAATGCGC TCCGCTGGG AGAGCCTAT GCAATGCTC
2101 AGAGAGAGC CATCAGCTC CTTCTGAGAG AAGCCATCAG AGTCCCTCTG
2151 AGAGAGACA TCACATCCC TCTGAGAGAA GCAATGCGC TCCCTCTGAG
2201 AGAGCATT GCAATGCTC TGAGAGAGA CATGCGATC CTTCTGAGAG
2251 AAGACATCAC ACTCCCTCAG AGAAGGCCA TCACATCCC TCTGAGAGAA
2301 GCAATCAGC TCCCTCTGAG AGAGAGCTC ACATGCTCTT GAGAGAGAG
2351 CTTCTCAGC TCTCTGAGAG GAGCATGCG ACTCCCTCTG AGAGAGATC
2401 TCACAGTCC TTTGAGAGGA GCAATGCTG GATTCTGAG AGAGTCAACA
2451 GTCTCTCAG GAGAGAGC CTGACCTCT TGGAGAGAG CCGTCTGAGT
2501 GCTCTGAGA GAGAGAGCA CAGTCTCTCT GCGAAGCTC CTACAGTCC
2551 CTTCTGAGAG AGCATTCGCA GTCCCTCTGG GATGAGGCAA GCGAGGAGCT
2601 CTGAGAGAG ACTGCGANT TCTCTGAGA GAAAGCTCA CAGTCTCTT
```

WO 01/12659

PCT/IB00/01496

```
2651 GAGATGAGGC CAGGAGGGCC CTCGGGAGG AACCATGCA GTCCCTCTGA
2701 GAGAGAGCGA CCGAGTCCCC TTAGAGAGG ACTCAAGTAC AGTTTCTCTG
2751 GAGAGAGGGC CAGGCTAGCT TTCTTTAGAG ATTTCAGCAA TCACACACT
2801 CTCCTCGGGA CCACACATAA AATCCCAA GCAGGGCAG TGTGGAGGCC
2851 TGAGGTACT CGATGAGGCC AGCTCCGCC CATTATCTA TTCTCTTAG
2901 TCTTCATCGT GGTGCGCTTT CAGGCTCTT TTCTCTCTCA GCCTCTGCT
2951 CCAATTCTGT CCGCCCGCAG GTGCAAGGC TTCTATTCT CTCTACCGGG
3001 GGGAGAGGGC CTGAGATGG CTCCTTATT TTCTTAGAT GAATAAGGG
3051 GCAGTTAATT AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA
```

#### BLAST Results

No BLAST result

#### Medline entries

No Medline entry

#### Peptide information for frame 2

ORF from 47 bp to 2863 bp; peptide length: 939  
Category: similarity to unknown protein  
Classification: unclassified  
Prosite motifs: ATP\_GTP\_A (824-832)

```
1 HEESDGGSD SCYRISDCH ELAPHYLGA WTDFSEQPOL LEDLQRLIAA
51 ELASDIPPO VPPPLASGLV LKTFICLQGG KCSGLWCHNK LQTSQPTLL
101 IYPOHLVRY PEHGCDFVLM LAFPLAIGR SQISRYEARD RPVIKRPIS
151 PGRMAKITT QAKSPFTET DUGSGPSOR APQWVIRG OBRMOLVEX
201 TETRAQMTZ FTQVHMLPES DSESTORER AKVNTARTSD SKYPMKITE
251 RLRFHREFTT MHTTICSPS RELAAHLRK RIGATOTSTA SLKQVAKPS
301 OPTFMOLLFO SLERATUTAR VILASVGRP VICTPFDMLA ASKPTTQCP
351 ADQYCLP331 KDKRSADKL TTAGSTIQE DILMGQTVOC RSKQPPRAY
401 STQPMPLALP APTDGGGIA TUTASVQQL RVVQDSESR SKAPFARET
451 SQQSKMLST KTRVQAKCR ILGSPVETK MHLKGLKLT HKENHPSFT
501 KERTYMGPSZ RTAKHPSMHN HSPSPSPQR SILERHNSP SQKACSPSR
551 EMLGSPERS HSPSPGHCN SPFERSKIEL RENGHLPSQ RHHGSPSPQ
601 HNSPERSHR SPERSHSPR SEKRNHSPQ RSHGSPERS HCSFPERNR
651 SPSPERSHOP SERHNSPSP RSHSPSPRS RSHSPSRSH SPBKHNSP
701 SERHNSPSE RNCSPERS RSPSPSRNR SPSPSRNSP SEKSHNSPSE
751 RSHNSPERR RSHFLERSH SLERHSPR SEKSHNSPSE RSHRISERS
801 HSPSESHLS FLERHNSPS ERGHSHSC TOLSPERSH HPSGHOCOP
851 TSESHNSSC ERTHNSPSEM RGRSPSPRN CSPERSERS FLKGLKYSF
901 PGRSPSHLS RDTKHQTTLL CTKHSPRAG QNRPPEAR
```

#### BLASTP hits

No BLASTP hits available

#### Alert BLASTP hits for DEF08es3\_0g11, frame 2

TRNDL:AF061185.1 gene: "car90"; product: "cyst germination specific acidic repeat protein precursor"; Phytophthora infestans cyst germination specific acidic repeat protein precursor (car90) gene, complete cds., M = 1, Score = 457, E = 2.3e-39

TRNDL:AC004361.38 gene: "F1492.41"; product: "putative proline-rich protein"; Arabidopsis thaliana chromosome II BAC F1492 genomic sequence, complete sequence., M = 1, Score = 340, E = 4.2e-27

TRNDL:AF062653.1 product: "plenty-of-proline-101"; Mus musculus plenty-of-proline-101 mRNA, complete cds., M = 1, Score = 313, E = 3.6e-24

PIR:P00099 son3 protein - human (fragment), M = 1, Score = 292, E = 1.2e-22

>TRNDL:AF061185.1 gene: "car90"; product: "cyst germination specific acidic repeat protein precursor"; Phytophthora infestans cyst germination specific acidic repeat protein precursor (car90) gene, complete cds.

Length = 1,489

## HSPs:

Score = 457 (68.6 bits), Expect = 2.3e-39, P = 2.3e-39  
Identities = 91/444 (20%), Positives = 239/444 (53%)

Query: 475 SPVKRTMIRHLKDLTKKHNPSTY-RETPRGPSERTRNHPSPWNRHNSPSEASQSSL 533  
+P + T + + + T + + + E T P + E T + P + +P+E + +S  
Sbjct: 584 APTEETNIAPIEET-TYAPTEETTYAPAEETPIEPEETTYAPTEETTYAPTEETTYAST 642  
Query: 534 ERNHSPPQRHNSPSEKHSPPSEKHSPPQRHNSPSEKHSPPSEKHSPPSEKHSPPSEKHSPP 593  
E + +P+ + +P+ + P+E + +P+ +P E + + +E +P+ + +  
Sbjct: 643 ECTTYAPTEETTYAPAEETPIEPEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTY 702  
Query: 594 RQPSQRHNSPSEKHSPPSEKHSPPSEKHSPPQRHNSPSEKHSPPSEKHSPPSEKHSPP 653  
P+ + +P+E + +P+E + +P+E +P + + GP+E + +P+E +P+  
Sbjct: 703 YAPAEETTYEPEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTE 762  
Query: 654 QRHNSPSEKHSPPSEKHSPPSEKHSPPSEKHSPPSEKHSPPSEKHSPPSEKHSPPSEKHS 713  
+ + P+E + P+ + P+ +P+E + +P+E + +P+E + +P+E + P+E +  
Sbjct: 763 ECTTYAPTEETTYEPEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTY 822  
Query: 714 CSPSEKHSPPSEKHSPPSEKHSPPSEKHSPPSEKHSPPSEKHSPPSEKHSPPSEKHS 773  
+P+E + P+E +P+E +P+E+ +P+E+ +P+E +P+E +P+E +P+  
Sbjct: 823 YAPTEETTYEPEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTE 882  
Query: 774 ERNHSPPSEKHSPPSEKHSPPSEKHSPPSEKHSPPSEKHSPPSEKHSPPSEKHSPPSEK 832  
E + +P+ + +E + + +P+E+ +P+E +P+E + + +T  
Sbjct: 883 ECTTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTY 942  
Query: 833 HSPSEKHSPPSEKHSPPSEKHSPPSEKHSPPSEKHSPPSEKHSPPSEKHSPPSEKHS 892  
+P+E + +P+ +P+E + +E + +E +P+E P+ +P+E + +P+  
Sbjct: 943 YAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTE 1002  
Query: 893 KEGKYSPPGERPSHLSGPKHOTT 918  
+E Y+ P E + + + + + T  
Sbjct: 1003 EE-TTYA-PEETTYAPAEETTYEPT 1026

Score = 445 (64.8 bits), Expect = 4.5e-38, P = 4.5e-38  
Identities = 83/394 (21%), Positives = 212/394 (53%)

Query: 502 ETPRGPSERTRNHPSPWNRHNSPSEKHSPPSEKHSPPSEKHSPPSEKHSPPSEKHSPPSEK 561  
E T P + E T + P + +E + +P+ + +P+ +P+E +  
Sbjct: 763 ECTTYAPTEETTYEPEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTY 822  
Query: 562 RQPSQRHNSPSEKHSPPSEKHSPPSEKHSPPQRHNSPSEKHSPPSEKHSPPSEKHSPP 621  
+P+ + P E + + +E +P+ + P+ +P+E + +P+E + P+  
Sbjct: 823 YAPTEETTYEPEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTE 882  
Query: 622 ERNHSPPQRHNSPSEKHSPPSEKHSPPSEKHSPPSEKHSPPSEKHSPPSEKHSPPSEKHS 681  
E + +P+ + P+E + +E +P+ + P+E + P+ + P+ +  
Sbjct: 883 ECTTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTY 942  
Query: 682 RQPSQRHNSPSEKHSPPSEKHSPPSEKHSPPSEKHSPPSEKHSPPSEKHSPPSEKHSPP 741  
+P+E + +P+E + +P+E +P+E + P+E + +P+E +P+E +P+  
Sbjct: 943 YAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTE 1002  
Query: 742 ERNHSPPSEKHSPPSEKHSPPSEKHSPPSEKHSPPSEKHSPPSEKHSPPSEKHSPPSEK 800  
E+ +P+E + +P+E + P E + +E + +P+E + +E + + +S +  
Sbjct: 1003 ECTTYAPTEETTYAPAEETTYEPEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTY 1062  
Query: 801 HSPSEKHSPPSEKHSPPSEKHSPPSEKHSPPSEKHSPPSEKHSPPSEKHSPPSEKHS 860  
+P+E+ + P E + +P+E + + +P+E + +P+ +E + +P+  
Sbjct: 1063 YAPAEETTYEPEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTE 1122  
Query: 861 ETPNPSPPQRHNSPSEKHSPPSEKHSPPSEKHSPPSEKHSPPSEKHSPPSEKHSPPSEK 894  
E T +P+E P+ +P+E + P+E  
Sbjct: 1123 ECTTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTY 1186

Score = 439 (65.9 bits), Expect = 2.0e-37, P = 2.0e-37  
Identities = 84/421 (20%), Positives = 223/421 (52%)

Query: 475 SPVKRTMIRHLKDLTKKHNPSTY-RETPRGPSERTRNHPSPWNRHNSPSEASQSSL 533  
+P + T + +E T+ + + E T P + E T + P + +P+E + +S  
Sbjct: 848 APTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTE 906  
Query: 534 ERNHSPPQRHNSPSEKHSPPSEKHSPPSEKHSPPSEKHSPPSEKHSPPSEKHSPPSEKHS 593  
E + +P+ + +P+ + P+E + +P+ +P E + + +E +P+ + +  
Sbjct: 907 ECTTYAPTEETTYAPAEETTYEPEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTY 966



969

Query: 801 KSPREKHLPLERKNCSPKRNHSSGKTCSPSPSRHSPSGHQGTSESRHSC 860  
+P+E+ +P E + +T +P+E + P+ +E +  
Sbjct: 719 YAPTEETTYAPTEETHYAPTEETTYGTEETTYAPTEETTYAPTEETTYAPTEETTYEPT 778

Query: 861 ERTNHSPSDHPGCPSPGRNHCSPSRKNSPLKGLATSPFGRPSHSLRDFKQTT 918  
T ++P+E P+ +P+E + +P +E Y P E +++ + + +T  
Sbjct: 778 GATTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYEPT+TYE+PTEETTYAPTEETTYEPT 834

Score = 428 (64.2 bits), Expect = 3.1e-36, P = 3.1e-36  
Identities = 89/140 (20%), Positives = 228/440 (51%)

Query: 473 KCSFNRTHHRLKDLTKEDHNPSTYR-ENTPGPSPERTNHPSHHNSPSPSRNS 531  
P P + T + R+ Y+ ++ E T P+E T + P+ P+E + +  
Sbjct: 470 PTEETTYAPTEETTYAPTEETTYASTETTYAPTEETTYAPTEETTYEPTETTYA 528

Query: 532 SLERNHSPDQNHCSPPSRNHCSPSPSRHSPSQNHCSPPSRHCSLSEGLKSPQR 591  
E +P+E + +P+ + +P+E + +P+ +P+E + +P+E + +P+E + +P+E +  
Sbjct: 529 PTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYEPTETTYAPTEETTYAPTEET 588

Query: 592 SHQPSQRNHCSPSRHNSPSPSRHNSPSPSRHNSPSPSRHNSPSPSRHNS 651  
+ P + ++P+E + +P+E + P+E +P++ + P+E + +E +  
Sbjct: 589 TNYAPTEETTYAPTEETTYAPTEETTYEPTETTYAPTEETTYAPTEETTYASTETTYA 648

Query: 652 PSQRNHCSPSRHNSPSPSRHNSPSPSRHNSPSPSRHNSPSPSRHNSPSPSRHNS 711  
P++ +P+E +P++ +P + +P+E + +P+E + +P+E + +P+E +  
Sbjct: 649 PTEETTYAPTEETTYEPTETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEET 708

Query: 712 SHCSPSRHCSPSRHNSPSPSRHNSPSPSRHNSPSPSRHNSPSPSRHNS 771  
+P+E + +P+E +P+E ++P E++ P+E + +P+E ++P E + +  
Sbjct: 708 TPEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYA 768

Query: 772 LLESHNSPSPSRHNSPSPSRHNSPSPSRHNSPSPSRHNSPSPSRHNS 830  
E + P+ ++ E + P +P+E++ +P E + P+E ++ +  
Sbjct: 769 PTEETTYEPTETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYEPTETTYAPTEET 828

Query: 831 TCHSPSRHNSPSPSRHNSPSPSRHNSPSPSRHNSPSPSRHNSPSPSRHNS 890  
T +P+E + +P+ +E + +E+ +P+E P+ +P+E + +  
Sbjct: 829 TPEETTYEPTETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYEPTETTYA 888

Query: 891 PLFLGLATSPFGRPSHSLRD 912  
P KE Y+ P E +++ +  
Sbjct: 889 PTKG-TTYA-PTEETTYASTEE 908

Score = 427 (64.1 bits), Expect = 4.0e-36, P = 4.0e-36  
Identities = 81/394 (20%), Positives = 213/394 (54%)

Query: 502 ETPPGSPERTNHPSHHNSPSPSRHNSPSPSRHNSPSPSRHNSPSPSRHNS 561  
E T CPSE T + P+ +P+E + E + P+ + +P+E +  
Sbjct: 739 EETTYGTEETTYAPTEETTYAPTEETTYAPTEETTYEPTETTYAPTEETTYAPTEETTY 798

Query: 562 RSPQRNHCSPSPSRHNSPSPSRHNSPSPSRHNSPSPSRHNSPSPSRHNS 621  
+P+ +P E + + +E ++P++ +P+ +P+E + +P+E + +P+  
Sbjct: 798 YAPTEETTYAPTEETTYEPTETTYAPTEETTYEPTETTYEPTETTYAPTEETTYAPTEETTYAP 858

Query: 622 EPNRSPQRNHCSPSPSRHNSPSPSRHNSPSPSRHNSPSPSRHNSPSPSRHNS 681  
E+ +P++ +P+E +P+E +P++ +P+E ++ + +P +  
Sbjct: 859 EATTYAPTEETTYAPTEETTYEPTETTYAPTEETTYAPTEETTYASTETTYAPTEETTY 918

Query: 682 RSPSRHNSPSPSRHNSPSPSRHNSPSPSRHNSPSPSRHNSPSPSRHNSPSPSRHNS 741  
+P+E + +P+E + +P+E +P+E + +P+E + +P+E +P+  
Sbjct: 919 YAPTEETTYEPTETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYEPT 978

Query: 742 EKSHNSPSPSRHNSPSPSRHNSPSPSRHNSPSPSRHNSPSPSRHNSPSPSRHNS 800  
E++ ++P+E + +P+E ++P+E ++ E + +P+E + E + + E +  
Sbjct: 978 EATTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYEPTETTYAPTEETTY 1038

Query: 801 KSPREKHLPLERKNCSPKRNHSSGKTCSPSPSRHSPSGHQGTSESRHSC 860  
+P+E++ + E + +P+E ++ +T +P+E + P+ +E + +  
Sbjct: 1039 YAPTEETTYASTETTYAPTEETTYAPTEETTYEPTETTYAPTEETTYAPTEETTYAPT 1098

Query: 861 ERTNHSPSDHPGCPSPGRNHCSPSRKNSPLK 894  
E T ++P+E P+ P+E + +P +E  
Sbjct: 1099 EATTYAPTEETTYAPTEETTYEPTETTYAPTEETTYAPTEETTYAPTEETTYAPTEET 1132

Score = 424 (63.6 bits), Expect = 8.9e-36, P = 8.9e-36  
Identities = 81/394 (20%), Positives = 210/394 (53%)

Query: 502 ETPPGSPERTNHPSHHNSPSPSRHNSPSPSRHNSPSPSRHNSPSPSRHNS 561  
E T P+E T + P+ +P+E + +E +P++ +P+ +P+E +  
Sbjct: 939 EATTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYEPTETTYAPTEETTYAPTEET 998

971

Subject: 676 KTYTAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTY 735  
Query: 850 RTSEKSHSSCERTANSPSNHNGPBGHNSPSEKSHSPLEGLKTSYFPCERPSHS 908  
E T E T ++P-E P+ -P-E + P-E T- P-E +++  
Subject: 736 APIETTYGTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTY 792  
Score = 420 (62.0 bits), Expect = 2.3e-35, P = 2.3e-35  
Identities = 82/393 (20%), Positives = 206/393 (52%)  
Query: 502 KTYPGPSETRHNPWNHNSPSEKSHSPSEKSHSPSEKSHSPSEKSHSPSEKSHSPSEKSHSPSEKSHSP 561  
E T- P-E T + P+ -P-E + -P-E ++P+ + P+ + P-E +  
Subject: 971 KTYVTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTY 1030  
Query: 562 NSPQSHNSPSEKSHSPSEKSHSPSEKSHSPSEKSHSPSEKSHSPSEKSHSPSEKSHSPSEKSHSP 621  
-P+ + P-E + -P-E ++P+ + P+ + P-E + -P-E + P+  
Subject: 1031 YAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTY 1090  
Query: 622 EKSHNSPQSHNSPSEKSHSPSEKSHSPSEKSHSPSEKSHSPSEKSHSPSEKSHSPSEKSHSPSEKSHSP 681  
E -P+ + P-E + -P-E P+ + P-E -P+ + P+  
Subject: 1091 KETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTY 1150  
Query: 682 NSPQSHNSPSEKSHSPSEKSHSPSEKSHSPSEKSHSPSEKSHSPSEKSHSPSEKSHSPSEKSHSP 741  
P-E + ++P-E + ++P-E -P-E + P+ -P-E -P-E ++P+  
Subject: 1151 YCPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTY 1210  
Query: 742 EKSHNSPSEKSHSPSEKSHSPSEKSHSPSEKSHSPSEKSHSPSEKSHSPSEKSHSPSEKSHSP 800  
E+ -P-E + + P-E ++P-E + E + -P-E + E + E  
Subject: 1211 KETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTY 1270  
Query: 801 NSPQSHNSPSEKSHSPSEKSHSPSEKSHSPSEKSHSPSEKSHSPSEKSHSPSEKSHSPSEKSHSP 860  
++P +++ P-E + -P-E ++ + T -P-E + P-E -E + +  
Subject: 1271 YAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTY 1330  
Query: 861 KTYHNSPQSHNSPSEKSHSPSEKSHSPSEKSHSPSEKSHSPSEKSHSPSEKSHSPSEKSHSPSEKSHSP 885  
E T ++P-E P P S C+ E  
Subject: 1331 KETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTY 1390  
Score = 419 (62.9 bits), Expect = 3.0e-35, P = 3.0e-35  
Identities = 83/411 (20%), Positives = 215/411 (52%)  
Query: 502 KTYPGPSETRHNPWNHNSPSEKSHSPSEKSHSPSEKSHSPSEKSHSPSEKSHSPSEKSHSPSEKSHSP 561  
E T P-E T + P+ -P-E + E ++P+ + P+ + P-E +  
Subject: 947 KETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTY 1006  
Query: 562 NSPQSHNSPSEKSHSPSEKSHSPSEKSHSPSEKSHSPSEKSHSPSEKSHSPSEKSHSPSEKSHSP 621  
-P+ + P-E + -P-E ++P+ + P+ + P-E + -P-E + P+  
Subject: 1007 YAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTY 1066  
Query: 622 EKSHNSPQSHNSPSEKSHSPSEKSHSPSEKSHSPSEKSHSPSEKSHSPSEKSHSPSEKSHSPSEKSHSP 681  
E -P+ + P-E + -P-E -P+ + P-E -P+ + P+  
Subject: 1067 KETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTY 1126  
Query: 682 NSPQSHNSPSEKSHSPSEKSHSPSEKSHSPSEKSHSPSEKSHSPSEKSHSPSEKSHSPSEKSHSP 741  
-P-E + ++P-E + ++P-E P-E + -P-E + -P-E -P-E P+  
Subject: 1127 YAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTY 1186  
Query: 742 EKSHNSPSEKSHSPSEKSHSPSEKSHSPSEKSHSPSEKSHSPSEKSHSPSEKSHSPSEKSHSP 800  
++ -P-E + -P-E ++P-E + E + P-E ++ E + + E +  
Subject: 1187 KETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTY 1246  
Query: 801 NSPQSHNSPSEKSHSPSEKSHSPSEKSHSPSEKSHSPSEKSHSPSEKSHSPSEKSHSPSEKSHSP 860  
++P-E + -P-E + -P-E ++ + T + P-E + P+ -E + +  
Subject: 1247 YAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTY 1306  
Query: 861 KTYHNSPQSHNSPSEKSHSPSEKSHSPSEKSHSPSEKSHSPSEKSHSPSEKSHSPSEKSHSPSEKSHSP 912  
E T + P+ P+ -P-E + -P-E T P-E + +P+ +  
Subject: 1307 KETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTY 1366  
Score = 415 (62.3 bits), Expect = 8.0e-35, P = 8.0e-35  
Identities = 84/423 (19%), Positives = 218/423 (51%)  
Query: 473 KCSFVXTHHRLKDLTKHNSPSEKSHSPSEKSHSPSEKSHSPSEKSHSPSEKSHSPSEKSHSPSEKSHSP 531  
P + T + E T- ++ E T P-E T + P+ -P-E + +  
Subject: 878 PTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTY 936  
Query: 532 NSPQSHNSPSEKSHSPSEKSHSPSEKSHSPSEKSHSPSEKSHSPSEKSHSPSEKSHSPSEKSHSP 591  
E ++P+ + P+ + -P-E + -P+ P-E + ++P+ ++P+  
Subject: 937 PTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTY 996  
Query: 592 NSPQSHNSPSEKSHSPSEKSHSPSEKSHSPSEKSHSPSEKSHSPSEKSHSPSEKSHSPSEKSHSP 651  
+ P+ ++P-E + -P-E + -P-E -P+ + P-E + -P-E +

Sbjct: 997 TWAPTEETTYAPTEETTYAPAEETPYETETTYAPTEETTYAPTEETTYASTETTYA 1054  
Query: 652 PSQRSHGSPERKSHSPERKSHSPAKRSHSPERKSHSPERKSHSPERKSHSPERKSHSPER 711  
P++ + P+E + P++ + P+ + P+G + P+G + P+G + P+G  
Sbjct: 1057 PTEETTYAPAEETPYETETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPAE 1116  
Query: 712 SHCSPEKSHSPERKSHSPERKSHSPERKSHSPERKSHSPERKSHSPERKSHSPER 771  
+ P+E + P+G + P+G + P+G + P+G + P+G + P+G + P+G + P+G + P+G + P+G + P+G  
Sbjct: 1117 PTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTY 1176  
Query: 772 LLERKSHSPERKSHSPERKSHSPERKSHSPERKSHSPERKSHSPERKSHSPERKSHSPER 830  
E + P++ + P++ + P++ + P++ + P++ + P++ + P++ + P++ + P++ + P++ + P++ + P++ + P++  
Sbjct: 1177 PTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTY 1236  
Query: 831 TCHSPERKSHSPERKSHSPERKSHSPERKSHSPERKSHSPERKSHSPERKSHSPERKSHSPER 890  
T + P+E + P+G + P+G + P+G + P+G + P+G + P+G + P+G + P+G + P+G + P+G + P+G  
Sbjct: 1237 PTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTY 1296  
Query: 891 PLE 694  
P +E  
Sbjct: 1297 PTE 1300  
Score = 403 (60.5 bits), Expect = 1.6e-33, P = 1.6e-33  
Identities = 84/394 (21%), Positives = 213/394 (54%)  
Query: 501 RERTPGPSEKTHWPKSHSPERKSHSPERKSHSPERKSHSPERKSHSPERKSHSPERKSHSPER 560  
PSE T + P +P+E +E + + + +P++ +P+E  
Sbjct: 319 RERTYAPAEETTYAPAEETTYAPTEETTY--DVEETTYVTEETTY--APTEETTYAPTEET 375  
Query: 561 WASPSQRSHGSPERKSHSPERKSHSPERKSHSPERKSHSPERKSHSPERKSHSPERKSHSPER 620  
+ + + C E + + +E +P++ + P++ +P+E + P+E + P+G + P+G  
Sbjct: 376 HYAHENP-COT-CVTHAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAP 433  
Query: 621 PERKSHSPERKSHSPERKSHSPERKSHSPERKSHSPERKSHSPERKSHSPERKSHSPERKSHSPER 680  
+E +P++ + P+E++ +P+E +P++ + P+E +P+E + P+G + P+G + P+G + P+G + P+G + P+G + P+G  
Sbjct: 434 PTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTY 493  
Query: 681 HNSPERKSHSPERKSHSPERKSHSPERKSHSPERKSHSPERKSHSPERKSHSPERKSHSPERKSHSPER 740  
+ +E + +P+E + +P+E + P+E + P+E + P+E + P+E + P+E + P+E + P+E + P+E + P+E + P+E  
Sbjct: 494 TYASTETTYAPTEETTYAPAEETPYETETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTY 553  
Query: 741 SEKSHSPERKSHSPERKSHSPERKSHSPERKSHSPERKSHSPERKSHSPERKSHSPERKSHSPER 799  
+E++ +P+E + P+E +P E + + + E + P E + + E + + E  
Sbjct: 554 TETTYAPAEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTY 613  
Query: 800 SHSPERKSHSPERKSHSPERKSHSPERKSHSPERKSHSPERKSHSPERKSHSPERKSHSPERKSHSPER 859  
+ P+E++ +P E + P+E +P+ T +P+E + P+ +E + +  
Sbjct: 614 PTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTY 673  
Query: 860 CERTANSPSQRSHGSPERKSHSPERKSHSPERKSHSPERKSHSPERKSHSPERKSHSPERKSHSPER 894  
E T +P+E P +P+E + P+G + P+G  
Sbjct: 674 TETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTY 708  
Score = 398 (59.7 bits), Expect = 5.5e-33, P = 5.5e-33  
Identities = 84/402 (20%), Positives = 209/402 (51%)  
Query: 475 SPVKTWPKSHSPERKSHSPERKSHSPERKSHSPERKSHSPERKSHSPERKSHSPERKSHSPERKSHSPER 533  
+P + T + + + T+ + E TP P+E T + P+ +P+E + +S  
Sbjct: 992 APTEETTYAPAEETTYAPTEETTYAPAEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTY 1050  
Query: 534 ERKSHSPERKSHSPERKSHSPERKSHSPERKSHSPERKSHSPERKSHSPERKSHSPERKSHSPERKSHSPER 593  
E +P++ + P++ + P+E + P++ +P E + + +E +P++ +  
Sbjct: 1051 ZETTYAPTEETTYAPAEETPYETETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTY 1110  
Query: 594 RGSQRSHGSPERKSHSPERKSHSPERKSHSPERKSHSPERKSHSPERKSHSPERKSHSPERKSHSPER 653  
P++ + P+E + P+E + P+E + P+G + P+G + P+G + P+G + P+G + P+G + P+G + P+G + P+G + P+G + P+G  
Sbjct: 1111 YAPAEETPYETETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTY 1170  
Query: 654 ORSHGSPERKSHSPERKSHSPERKSHSPERKSHSPERKSHSPERKSHSPERKSHSPERKSHSPERKSHSPER 713  
+ + P+E + P+ + P+ + P+E + P+E + P+E + P+E + P+E + P+E + P+E + P+E + P+E + P+E + P+E + P+E  
Sbjct: 1171 RERTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTY 1230  
Query: 714 CSPERKSHSPERKSHSPERKSHSPERKSHSPERKSHSPERKSHSPERKSHSPERKSHSPERKSHSPER 773  
+P+E + P+E +P+E +P+E++ +P+E + +P + P E + +  
Sbjct: 1231 YAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTY 1290  
Query: 774 ERKSHSPERKSHSPERKSHSPERKSHSPERKSHSPERKSHSPERKSHSPERKSHSPERKSHSPERKSHSPER 833  
E + +P+E + E E + P+ + +P E + +P+E + + +T +  
Sbjct: 1291 KATTYAPTEETTYAPTE-----ETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTY 1343  
Query: 834 SPERKSHSPERKSHSPERKSHSPERKSHSPERKSHSPERKSHSPERKSHSPERKSHSPERKSHSPER 896

974

Report for DKFZphtes3\_8g11.2

975

PCT/IB00/01496

```

      From: for DKFzptes3_gll.2
PS00017    639->847  ATP_GTP_A      PC0C00017
(No Prim data available for DKFzptes3_gll.2)

```



WO 01/12659

PCT/IB00/01496

DEFPntes3\_Bp5  
-----

group: testes derived

DEFPntes3\_Bp5 encodes a novel 166 amino acid protein nearly identical to human KIAA087 protein.

The novel protein is a new splice variant of KIAA087.  
No informative BLAST results; No predictive prosite, pfam or SCOP motifs.

The new protein can find application in studying the expression profile of testis-specific genes.

KIAA087, alternative spliced

complete cDNA, complete cds, EST hits

Sequenced by MediGenomix

Locus: unknown

Insert length: 2762 bp  
No poly A stretch found, no polyadenylation signal found

```

1  CCACATCGG  CGGTGTCTCC  AGCAGCTGCC  GCGCGGTGCG  CGAGCTGTGC
51  CAGAGCAGCG  GGAAGGTCTG  GAAGGAGCGG  TTCCGGGTGA  GGTGACCTTC
101  CCTTATGAAA  CACTACAGCC  CAGCCGACTA  CGTCAATTGG  TTGGAAGAGT
151  ATAACTCTGG  GCAAAAGAGT  GGGTTAGAGG  GCGGGAAGAT  TGTAGCTCTG
201  TTCTCAAGAA  GGTCTCTTTC  AGAGCAGCTT  CCTGTGATAT  GCTTCAGTGA
251  CATTGAGAGC  CTTCAGGAGC  CACAGATTTT  TTTCAGAGAT  GAACGTGGTT
301  GTATCTCTAA  TATGGAAGGA  AGAAAGCTTT  TGACCTGGAA  ATACTACCCA
351  AAAAAAATTT  TTACTACTCT  GCGGCAAGAG  AGAGTCTTAA  ATATCTTTAA
401  GGGCTTTCTT  CAGCAGCAG  ATGACTATGA  GTGCTATCTT  GAAGGTCTTG
451  TATATATTTA  CAGTACTCTG  AATCTCTCTT  CCGCATCAGC  GCTCAAGAGC
501  ATCCAGGCGC  AATTATGAGC  CATCTGGAGC  CTCTCTTGCA  AAACCTCTGG
551  GGGCATTAAC  AGTCCGACCC  CAGGCTTGCG  CTCAAGGCGA  GGTGAATCAT
601  GCATATATAT  GGAATAGAGA  CTCAGAGGCG  AGGTGCTGGA  TGCCATGAAC
651  TATGTCTCTT  AGCAGCAACT  GAAGTTCAGG  GCGAATCGAA  TGCATTACTA
701  TAAATGCCCT  AACTTATATA  TGCATCAGGT  TTGATATCTG  AGAAGAGGAA
751  TCCCAATAGC  CATCTCTCTG  CTTTATTCTG  CAATTCTCTG  CGACTGTGGA
801  GTCCCACTGG  AGCCTGTCAA  CTTCCCAAGT  CACTTCTTAT  TAAGTGTGTT
851  CGAGAGGCGA  GAAGGGGCGA  CCGTGGACAT  CTTTGACTAC  ATCTATATAG
901  ATGCTTTTGG  GAAGGCGAAG  CAGCTCAGAG  TGAAGAATGT  CGAGTACTTG
951  ATCGGCGAGC  AGCTACTCTG  AGCACTGTAT  GGGGTGTCTA  ATCTCAAGAA
1001  GGTCTTACAG  AGATCTGTGG  GAAACTCTTT  AAGCTCTGGG  AAGCGGAAG
1051  GCATCGACCA  GTCATACGAG  CTCCTGAGAG  ACTGCTGGA  TCTCTATCTG
1101  GCATGTAGC  GCGACAGAGT  GCAATCTCTC  CTCCTGAGAG  CGAGGCTTTA
1151  CTTCCAGCTG  GGAATCTGGC  CAGAGAAGTC  TTCTCTCTTT  GTTTTGAGGG
1201  TGTCTGACAT  CTTCCAGCAC  ATCCAAATCC  TAGACCGGGG  GCAGCAGGGG
1251  GCGTGGGCTT  ACCTGTGTGA  GCAACTCTTA  GAGACATATG  AGCCAAAGAA
1301  GGAGGAGGTG  GGCCTAGAGG  TGAAGCTGCG  CTCGATGAG  AAGCACAGAG
1351  ATGTGTGCTA  CTGATCTGGG  CTCATTATAG  AGCATAGAG  GTATGCTAT
1401  AACTGTGTGA  TGTAGGCTCT  GGAACCCAGC  TGCATGATGG  GACACAGCTG
1451  GATCCGGAGC  ATGAACTCTC  AAGGCTCTCC  GCAAGGAGAC  CACAGACTTT
1501  TCTATATAGT  CTTGTGTGAG  GAGGCTCTCT  GTGGATAGCC  AGCCGAGAAA
1551  AACTGTGAAT  ATAGCTGTGA  GCTTCAGAAA  ATCTACATCC  CTGACGTGGG
1601  AGCTTATCTT  TCGAGCTTTA  CTGGCACTCT  CTACATCTGA  AAGCGAGAGC
1651  TCGAGATCTG  GTATCCAGAA  GATCTGGAGT  TTGTCTATGA  AACGTGTGAG
1701  ATATATTACA  GTCCAGAGAA  AGAGACATTA  GATGATTAAT  GTGTGAGAG
1751  GACATTGAC  CTTTGTCTCT  GCTGCTATCT  TCCAGAGAAA  CGGAGCTCTG
1801  GAAGAAGAGC  TCTCCAGGGA  GCGCTCGGGA  CTTCTGTGAG  CAGGAGAGCC
1851  ACTCCAGAGC  TACTGTCTGT  TCTCTCTATC  TAACTTTAAA  TACCTGTGCG
1901  TCTTCCCGAG  CTCAGAGAGC  AATCTGTCTC  TCCGCTTACA  CTAGTGAATT
1951  AATCTGAAAG  GCATCTGTCT  ACTGGAGTGG  CTTGTATGCT  TGTCTGTGG
2001  TGCAGCTTTG  TGACATCTTG  TCTTCTATAG  GTCTCAGAGT  CGAGCTCTCT
2051  GTAATCATTC  TTGTATTTCA  CTCATCTGCC  CTCTCTCTCT  GCATTCTCTT
2101  CAGACATCTT  CTTCTGCTGG  ACAGATGGGG  TTATCATCTT  GCATATATTT
2151  CTTCTGTGAT  TCTCTGTGGA  AGCTGTGTCG  TCCGAGTGA  GAGCTATGTT
2201  TCTTTTACAC  CTGATTTTAG  TTGCTATAT  AGAGGTAAAG  TTGTGTCTTA
2251  TCTTGGCAGC  ATCTTAGAGA  TGGAGACATT  AAGAGCTTAA  TGTATTATAG
2301  AATCATTTGA  ATTTATTTTT  TCTTAATATG  TGAACACAGC  ATTTCAAGTG
2351  TTTATCTCTT  TTTTCTTTTA  ATTTTAATAT  GGAATATAGC  ACAGTTCTTC
2401  CTTCCATATT  CTTCTCTTGA  GTTTATGAGC  ATCTCTATTA  ATCATTAGTT
2451  TTCTATTTTA  TTACATTAAG  TCTTTTGA  AATTCGAAAT  ACTGAGCTTT
2501  GTCAATGGA  TTTTCCATAC  TCATCTAGAA  TTCTCTCAT  TTAATGACTT
2551  ACTTTTATT  TTTATTTTAA  AAAATCTACT  TCAATATCAT  CAGTAGCTCT
2601  TACATCAGTG  ATGGTCTCT  TTTGTAGTGA  GACATCAAAA  TCTATCTTA

```

WO 01/12659

PCT/IB00/01496

2651 ATGTTTGGCT TTAGAAGTCA TACTCATGG TCCTCAAGA CCAAAAAATG  
2701 AGGTTTGGCT TTCTAATCA GCAAAAAAA AATTAAATGA CTTAAAAA  
2751 AAAAAAAA GG

BLAST Results

No BLAST result

Medline entries

No Medline entry

Peptide information for frame 3

ORF from 105 bp to 1736 bp; peptide length: 544  
Category: known protein  
Classification: unclassified

1 MKHYSPTDYN MHLEETVRQ KAGLEAKIV ASFRKFFIE HVPKGFSDI  
51 EHLGSEPIFF EDELVCILMH KGSALTWET TAKSILYTL QDTLHRLA  
101 FLOQPDYES YLEGAVTIQ YCHPLSDISL KDIQAQDSI VELVCKTLNG  
151 HSRNPLAF RAGESHINE ELQSQVLA MHVLYDQL FGHMNYTH  
201 ALMLHNDML IRTGPIISM ELLTITAND LQVLEPNF PSHFLMNC  
251 GAGGATLDI FDIYIDAFGR GQLTYKCE YLIGHVTAA LYGVWVREY  
301 LQNVGHLLS LGRREGIDQ YQLARDELQ YLAMPYDQV LLIGARLYF  
351 HLGIMPEKSF CLVLKVLDI LQNTOLQDQ NGAVGYLQV TLHIERKRE  
401 EVDVVELAS DEANQPCIS LLLINAKRE CHWYLYQD PTCHKHDI  
451 KNDHNSLPM QKQPTFYVL VEGGKRYTA QEMLEYNVE QETSPDQCR  
501 YPSEFTGTH IPRALEIRY PEDLEPYET VQMLYAKKE HIDE

BLAST Hits

No BLAST hits available

Alert BLAST hits for ORFspites\_8q5, frame 3

TRENHLMEN:AB020682.1 gene: "XIAA0875"; product: "XIAA0875 protein";  
Homo sapiens mRNA for XIAA0875 protein, partial cds.; M = 1. Score =  
2832, E = 5.5e-295

>TRENHLMEN:AB020682.1 gene: "XIAA0875"; product: "XIAA0875 protein"; Homo  
sapiens mRNA for XIAA0875 protein, partial cds.  
Length = 621

HSPs:

Score = 2832 (424.9 bits), Expect = 5.5e-295, E = 5.5e-295  
Identities = 537/544 (98%), Positives = 537/544 (98%)

Query: 1 MKHYSPTDYNMHLEETVRQKAGLEAKIVASFRKFFIEHVPKGFSDIDLEGLPIFF 60  
MKHYSPTDYNMHLEETVRQKAGLEAKIVASFRKFFIEHVPKGFSDIDLEGLPIFF  
Subject: 85 MKHYSPTDYNMHLEETVRQKAGLEAKIVASFRKFFIEHVPKGFSDIDLEGLPIFF 144  
Query: 61 EDELVCILMHKGSALTWETAKSILYTLQDTHLHKAFLQPDYESYLEGAVTIQ 120  
EDELVCILMHKGSALTWETAKSILYTLQDTHLHKAFLQPDYESYLEGAVTIQ  
Subject: 145 EDELVCILMHKGSALTWETAKSILYTLQDTHLHKAFLQPDYESYLEGAVTIQ 204  
Query: 121 YCHPLSDISLKDIAQDSIVELVCKTLNGHSRNPSLAFRAGESHINEELQSQVLA 180  
YCHPLSDISLKDIAQDSIVELVCKTLNGHSRNPSLAFRAGESHINEELQSQVLA  
Subject: 205 YCHPLSDISLKDIAQDSIVELVCKTLNGHSRNPSLAFRAGESHINEELQSQVLA 264  
Query: 181 MHVLYDQLFGRHMYTHALMTHQVLIIRGTGPISSHLLYLTIANQLGVLEPNF 240  
MHVLYDQLFGRHMYTHALMTHQVLIIRGTGPISSHLLYLTIANQLGVLEPNF  
Subject: 245 MHVLYDQLFGRHMYTHALMTHQVLIIRGTGPISSHLLYLTIANQLGVLEPNF 324  
Query: 241 PSHFLMNCGAGGATLDIFDIYIDAFGRGQLTYKCEYLIGHVTAAALYGVWVREY 300  
PSHFLMNCGAGGATLDIFDIYIDAFGRGQLTYKCEYLIGHVTAAALYGVWVREY  
Subject: 325 PSHFLMNCGAGGATLDIFDIYIDAFGRGQLTYKCEYLIGHVTAAALYGVWVREY 384  
Query: 301 LQNVGHLLS LGRREGIDQ YQLARDELQ YLAMPYDQV LLIGARLYF LIGIMPEK 360  
LQNVGHLLS LGRREGIDQ YQLARDELQ YLAMPYDQV LLIGARLYF LIGIMPEK

```
Subject: 385 LQHWGMLSLGRKEDIGSTOLLADSLDLYLAMYPOQVGLLLQALYPLUGIWPES-- 442
Query: 361 CLVLAFLDIQHTLDPQNGAVGLVQNTLEHIERKKEVGVVELASDEKHQVCTIS 420
      VLDLQHTLQTLDPQNGAVGLVQNTLEHIERKKEVGVVELASDEKHQVCTIS
Subject: 443 -----VLDLQHTLQTLDPQNGAVGLVQNTLEHIERKKEVGVVELASDEKHQVCTIS 497
Query: 421 IGLIMENKRYTMCVLYCHQPTCMGHEWIRMMQVSLPHGHQPFYVVLVEGSCRTAA 480
      IGLIMENKRYTMCVLYCHQPTCMGHEWIRMMQVSLPHGHQPFYVVLVEGSCRTAA
Subject: 498 IGLIMENKRYTMCVLYCHQPTCMGHEWIRMMQVSLPHGHQPFYVVLVEGSCRTAA 557
Query: 481 QKLEYHVEPQEISHFVGRTFSEPTGTHYIPMAELEIRYPEDELPVYETVQHTSAKKE 540
      QKLEYHVEPQEISHFVGRTFSEPTGTHYIPMAELEIRYPEDELPVYETVQHTSAKKE
Subject: 558 QKLEYHVEPQEISHFVGRTFSEPTGTHYIPMAELEIRYPEDELPVYETVQHTSAKKE 617
Query: 541 HIDE 544
      HIDE
Subject: 618 HIDE 421
```

Pedant information for DRFiphea3\_Rq5, frame 3

Report for DRFiphea3\_Rq5.3

```
[LENGTH] 544
[MW] 63307.22
[pI] 5.82
[CDOML] YIPDRL:AB020682.1 gene: "KIA0875"; product: "KIA0875 protein"; Homo sapiens
name for KIA0875 protein, partial cds. 0.0
[FW] Alpha Beta
[FW] LOW_COMPLEXITY 1.84 s

SEQ HNRHSTPTQVWVLEEVEYQKAGLEAKIVASFSKRPFSYVPCNGSDIENLEGPCITF
SEG .....
PRD ccccccccccccccccccccccccccccccccccccccccccccccccccccccccc

SEQ EDELVCILMNEGKALTWYFAKILYLAQQLIMWLKAPLQKQRYELEGAVTIQD
SEG .....
PRD ccccccccccccccccccccccccccccccccccccccccccccccccccccccccc

SEQ YCPLSDIQLKDIQAGISIVELVCHTLAGIWRPPLAFAGESSHMEILQSQVLD
SEG .....
PRD ccccccccccccccccccccccccccccccccccccccccccccccccccccccccc

SEQ HNVLYDQLKFGKMDYMAKLTHQVLIARTGIPISMSLLYLTIAQLGVLPVPVF
SEG .....
PRD hhhhhcccccccccccccccccccccccccccccccccccccccccccccccccccc

SEQ PSHFLRWQCGAGATLQIDYIIDAIRKGLTVKCEYLIGHVYALYGVVWVRY
SEG .....
PRD ccccccccccccccccccccccccccccccccccccccccccccccccccccccccc

SEQ LQHWGMLSLGRKEDIGSTOLLADSLDLYLAMYPOQVGLLLQALYPLUGIWPESF
SEG .....
PRD hhhhhcccccccccccccccccccccccccccccccccccccccccccccccccccc

SEQ CLVLAFLDIQHTLDPQNGAVGLVQNTLEHIERKKEVGVVELASDEKHQVCTIS
SEG .....
PRD kkkkkkkkkkkkkkkkkkkkkkkkkkkkkkkkkkkkkkkkkkkkkkkkkkkkkkk

SEQ IGLIMENKRYTMCVLYCHQPTCMGHEWIRMMQVSLPHGHQPFYVVLVEGSCRTAA
SEG .....
PRD ccccccccccccccccccccccccccccccccccccccccccccccccccccccccc

SEQ QKLEYHVEPQEISHFVGRTFSEPTGTHYIPMAELEIRYPEDELPVYETVQHTSAKKE
SEG .....
PRD hhhhhhhcccccccccccccccccccccccccccccccccccccccccccccccc
```

SEQ HIDE

SEG ....

PRD CCCC

(No Prosite data available for DRFiphea3\_Rq5.3)

(No Pfam data available for DRFiphea3\_Rq5.3)

WO 01/12659

PCT/IB00/01496

DEFspites3\_0ml0  
\*\*\*\*\*

group: nucleic acid management

DEFspites3\_0ml0 encodes a novel 221 amino acid protein with strong similarity to polyadenylate-binding proteins.

The poly(A)-binding protein (PABP) binds to the messenger (mRNA) 3'-poly(A) tail found on most eukaryotic mRNAs and together with the poly(A) tail has been implicated in governing the stability and the translation of mRNA.

The new protein can find application in modulation of mRNA translation and processing/stability.

strong similarity to polyadenylate-binding protein

frame shift at Rp 707-710

Sequenced by MediGenomix

Locus: unknown

Insert length: 2107 bp

Poly A stretch at pos. 2052, polyadenylation signal at pos. 2033

```

1 CGGAAAGGTC GCGGCTTGTG TGCCTGGGG CAGCCTGCG GAGAAAGAAC
31 CCGACAGACC CCGATGACCC AAGGAGCTGG CTCCTAGTGG GGGAGCTCCA
101 CCGCAGAGTC ACTGAGGCGA TGCCTACGA GAAGTTCAAG CCGGAGGGCG
151 CATCTCTCTC CATCGGATC TCGAGGACTY TGATACCCAG GGGCTCTCTC
201 AATCTAGGCT ATCTGAACTT CCGAGATAGC AAGACAGCGC AGCATCTCTG
251 GAGACCACTG AATTTCGATG TTATAAGGG CAGGCACTA CGCATCATGT
301 GGTCTAGGCG TGATGATCA CTCGAGAAA GTGGAGTGGC CAGATATATG
351 GTTAAAGATC TGGATAGTCT CATTAATAT AAGGACTCTY ATGATACAGT
401 TCTGCTTTT GTTAACTCC TTCTGTSTAA GGTGCTTGT GATGAAATG
451 GTTCAAGGCG TTATGATTT GTACACTTTC AGACAGCGCA AGCAGTTTAA
501 AGAGCTATTA AAAAAATGAA CGAATGCTC CTAAATGCTC GCAAGATATT
551 TGTGGACAAA TTAAATCTTC CTAAAGAGCG AGAAGTCTCA CTGGAGCTA
601 GGGCAAGAGA GTTCCCAAT GTTACATCA AGAATTTGG AGAAGCATG
651 GATGATGAGC GCTTTAAGA TCTTTTGGC AGCTTGGAG CCGCTTTAG
701 TGTGAATTA TGCAGCTCA AGCTGAGAAA TCGAAGAT TTGATTTCT
751 AGGCTTTGAA AGGATGAG AGTCACAGAA AGCTTAGAT GAGATGAATG
801 GAAGGAGCTT CAGTGAAGA CAATTTTTC TTGCTGAGC TCGAAGAAA
851 GTGAGAGGCG AGACGAACT TAAGGACCA TTTGACAGA TGAGGACAGA
901 TAZGATCACC AGATACAGG TTCTTATCT TTATGAGA AATCTGATG
951 ATGTATTGA TGATGAGCT CTGGGAAGC GCTTTCTTC ATTTGATCA
1001 ATCACTAGTG CAAGATTTAT GATGAGAGT GTTGGAGCA AAGGCTTTGG
1051 TTTTGTATGT TTTCTCTGCC CAGAGAGGCG CACTAAGCA CTACAGAAA
1101 TGAGCGTAA AATTGTGGCC ACAAGGCAAT TGTATGAGC TTTAGCTCAG
1151 GCGAAGAAAG AGGCGAGGCG TTAGCTACT AGGAGTATA TGGAGAGAT
1201 GCGAGTGTG CAGCTGTGCG CCAAGCAGCG AGCAGCTCTY TCAGGTTACT
1251 TATAGCAGC TGTCCACAG ACTGAGAGCC ATGCTGCATA CTATCTCTCT
1301 AGGCAANTTG CTGCACTTAG ACAAGTCTC CCGTGAGTCT CTAGGTTGCG
1351 CAGACCTCAT GCTATTCAAA ATAGGCCAG TCGTATCCGC CCGAGTCTCT
1401 CTAGATGCC ATTTAGTACT ATCAGAGCAG CTCTCTCACA GOTTTCAGAA
1451 GTCATGTCAA CCGAGCTGTG TGCTAACCA TCAGACAGA CAGTGGTCTC
1501 AGCTCTGCA GCTGCTGCTG CTGCTGAGC TAGCCTCTCT GTGGGAGCG
1551 TCCACAGTGA TAAATATCT CCGGAGCTTC GCAATCTCA CCAAGATCT
1601 ATGACAGCAG CACAAGTTAC AATGACAGAG CTCTGCTGTC ATGTACAGG
1651 TCGAGAACTT TGACTCTCT CAGGTTGGC ATCTGGGCTC GCTTAAGAGC
1701 AAGAGCAAT GTTAGTGAA CGGCTCTTTC CTCTTATCA AGCAGTCAC
1751 CTAATCTCTG CTGGGAATAT CACTGCAATG TCTTGAGAG TTGATATTC
1801 AGAATCTCTT TATATCTCTG AGCTTCAGCA CTCACTCTCT TTAGAGTTG
1851 ATGAGGCTGT AGCTGTACTA CAGGCCAGC AGCTTAAGA GGTATCCAG
1901 AAGACAGTGA AGATGCTTAC CGCTGTCTCA ACTGTTTAA ATTATGACA
1951 GACACAGAAA AGAATTTGT GCTTACCGA AGAATAATAT CTAAACATG
2001 AGAATCTAT GGAAGAAAAA TTGCAAAATG TAAATATAA TATCCAAAT
2051 CTAAATATA AAAAAAAA AAAAAAAA AAAAAAAA
2101 AAAAAAG

```

BLAST Results  
\*\*\*\*\*

Entry NP01458 from database DBL:  
Human mRNA for polyA binding protein  
Score = 3420, E = 0.0e+00, identities = 1162/1243

## Medline entries

No Medline entry

## Peptide information for frame 2

ORF from 707 bp to 1936 bp; peptide length: 410  
Category: strong similarity to known protein  
Classification: unmet  
Protein motifs: RNP\_1 (10-18)  
RNP\_1 (112-120)

1 LMTDESKKSE GFGVSEFERN EDAQKAVDCH NGKELNGFOI YVGAQKIVE  
51 RYTELKRTPE QHEQORITPY QVHLVYVNL DQGLDEALK KAFSPETIT  
101 SKAVMHGGR SKGFGVCFB SPEEATKAVT DQGRIVATK FLVVALAGE  
151 EERQAVLME YMGNAVRA YMGAPPSG YMTAVPQTG MHAATFFSQ  
201 IARLRSPRM TAQGRAPFP QHEPALKRG APRVPESTR PASSQVPRN  
251 STORVHST OTVGRFPAK AHAATPAVE YTPETKAC YHPCQIRNA  
301 QPQVTHOLA VHVQOETLT ASKLASAPQ KQKMLGERL FFLQAMRPT  
351 LAGITONLL EIONSELLH LESPELRSK VQKAVVQK HQKQATQSA  
401 VMSATQVTV

## BLASTP hits

No BLASTP hits available

Alert: BLASTP hits for ORF2phex3\_5m10, frame 2

PIR:DHUHA polyadenylate-binding protein - human, M = 1, Score = 1931,  
P = 1.7e-199

PIR:146718 poly(A) binding protein - mouse, M = 1, Score = 1928, P =  
3.6e-198

>PIR:DHUHA polyadenylate-binding protein - human  
Length = 633

## HSPs:

Score = 1931 (289.7 bits), Expect = 1.7e-199, P = 1.7e-199  
Identities = 384/415 (92%), Positives = 394/415 (95%)

Query: 1 LMTDESKKSE GFGVSEFERN EDAQKAVDCH NGKELNGFOI YVGAQKIVE RYTELKRTPE 60  
Sbjct: 219 VHTDESKKSGFGVSEFERN EDAQKAVDCH NGKELNGFOI YVGAQKIVE RYTELKRTPE 278

Query: 61 QHEQORITPY QVHLVYVNL DQGLDEALK KAFSPETIT SKAVMHGGR SKGFGVCFB 120  
Sbjct: 279 QHEQORITPY QVHLVYVNL DQGLDEALK KAFSPETIT SKAVMHGGR SKGFGVCFB 338

Query: 121 SPEEATKAVT DQGRIVATK FLVVALAGE EERQAVLME YMGNAVRA YMGAPPSG YMTAVPQTG 174  
Sbjct: 339 SPEEATKAVT DQGRIVATK FLVVALAGE EERQAVLME YMGNAVRA YMGAPPSG YMTAVPQTG 398

Query: 175 RAPPSCGYMTAVPQTG MHAATFFSQ IARLRSPRM TAQGRAPFP QHEPALKRG APRVPESTR 234  
Sbjct: 399 RAPPSCGYMTAVPQTG MHAATFFSQ IARLRSPRM TAQGRAPFP QHEPALKRG APRVPESTR 458

Query: 235 PESTNRPASSQVPRN STORVHST OTVGRFPAK AHAATPAVE YTPETKAC YHPCQIRNA 294  
Sbjct: 459 PESTNRPASSQVPRN STORVHST OTVGRFPAK AHAATPAVE YTPETKAC YHPCQIRNA 517

Query: 295 QPQVTHOLA VHVQOETLT ASKLASAPQ KQKMLGERL FFLQAMRPT LAGITONLL EIONSELLH 354  
Sbjct: 519 QPQVTHOLA VHVQOETLT ASKLASAPQ KQKMLGERL FFLQAMRPT LAGITONLL EIONSELLH 577

Query: 355 ITGKLLIDWELLHLESPELRSK VQKAVVQK HQKQATQSA VMSATQVTV 410  
Sbjct: 579 ITGKLLIDWELLHLESPELRSK VQKAVVQK HQKQATQSA VMSATQVTV 633

Score = 315 (47.3 bits), Expect = 1.9e-27, P = 1.9e-27

**HSPs:**

983

[FUNGAT] 04.05.05 mRNA processing (5'-end, 3'-end processing and mRNA degradation) [S. cerevisiae, YER165w] 1a-54  
[FUNGAT] 30.03 organization of cytoplasm [S. cerevisiae, YER165w] 1a-54  
[FUNGAT] 30.10 nuclear organization [S. cerevisiae, YER165w] 1a-54  
[FUNGAT] 05.04 translation (initiation, elongation and termination) [S. cerevisiae, YER165w] 1a-54  
[FUNGAT] 04.05.99 other mRNA-transcription activities [S. cerevisiae, YER165w] 1a-15  
[FUNGAT] 11.01 stress response [S. cerevisiae, YER165w] 1a-12  
[FUNGAT] 04.01.04 RNA processing [S. cerevisiae, YER165w] 1a-12  
[FUNGAT] 04.99 other transcription activities [S. cerevisiae, YER165w] 1a-09  
[FUNGAT] 94 classification not yet clear-cut [S. cerevisiae, YER165w] 1a-08  
[FUNGAT] 03.19 recombination and DNA repair [S. cerevisiae, YER165w] 1a-07  
[FUNGAT] 03.13 meiosis [S. cerevisiae, YER165w] 1a-07  
[FUNGAT] 04.05.03 mRNA processing (splicing) [S. cerevisiae, YER165w] 1a-07  
[FUNGAT] 04.07 RNA transport [S. cerevisiae, YER165w] 1a-07  
[FUNGAT] 30.13 organization of chromosome structure [S. cerevisiae, YER165w] 1a-06  
[FUNGAT] 99 unclassified proteins [S. cerevisiae, YER165w] 1a-06  
[FUNGAT] 06.04 protein targeting, sorting and translocation [S. cerevisiae, YER165w] 1a-05  
[FUNGAT] 08.01 nuclear transport [S. cerevisiae, YER165w] 1a-05  
[FUNGAT] 11.04 DNA repair (direct repair, base excision repair and nucleotide excision repair) [S. cerevisiae, YER165w] 1a-05  
[FUNGAT] 03.01 cell growth [S. cerevisiae, YER165w] 1a-04  
[BL000308] Eukaryotic RNA-binding region RNP-1 proteins  
[SCOP] dist\_ 4.34.7.1.3 Sea-lethal protein (Drosophila melanogaster) 1a-17  
[PIRKM] nucleus 0.0  
[PIRKM] duplication 0.0  
[PIRKM] RNA binding 0.0  
[PIRKM] nucleolus 2a-09  
[PIRKM] tandem repeat 2a-09  
[PIRKM] single-stranded DNA binding 3a-06  
[PIRKM] DNA binding 3a-13  
[PIRKM] phosphoprotein 4a-10  
[PIRKM] ribosome 3a-08  
[PIRKM] mitochondrion 3a-08  
[PIRKM] alternative splicing 9a-11  
[PIRKM] chloroplast 2a-19  
[PIRKM] transcription regulation 2a-07  
[PIRKM] protein biosynthesis 3a-08  
[SUPTAM] nucleolin 6a-10  
[SUPTAM] glycine-rich RNA-binding protein 2a-07  
[SUPTAM] unassigned ribonucleoprotein repeat-containing proteins 2a-19  
[SUPTAM] polyomylate-binding protein 0.0  
[SUPTAM] ribonucleoprotein repeat homology 0.0  
[PROSITE] RNP\_1 2  
[PFAM] RNA recognition motif. (aka RRM, RBD, or RNP domain)  
[SWI] 17  
[SWI] 3D  
[SWI] LOW\_COMPLEXITY 5.42 1

SEQ HTDSESGSGPGFVPERHEDQKAYDHWGHELMKQIYVRAQKHYKTELENTTQ  
SEG .....  
1a1- .....  
SEQ HQQRITATYQVWLVKVLGQIDGALWAFSPGCTTSARVWEGGRSGPGVCFSS  
SEG .....  
1a1- .....CEEECCCTTTTNNNNHHHTTTTCCCECECTTCTTTECECTT  
SEQ FEETATVTDHGRIVATKPLVALAQKKEEQATLTNEYQNASVRAVTHQAPPSY  
SEG .....  
1a1- .....HHHHHHHHHTTTTCCCECECTTCTTTECECTT  
SEQ DWTAVPTQWHAATYFPGQILRLSPFMTAGGARHPFQWPSAIRPCAPVFTNRP  
SEG .....  
1a1- .....  
SEQ ASSQVPRVNSTQVARTSTQVGPFAAAAAAATPAVTVPRVTAAGVWMPQRMQAQ  
SEG .....  
1a1- .....  
SEQ PVTNQLAVNYGQETLTASRLAAPPKQKQHLGALFPLTQNHPTLAKRTIGALLE  
SEG .....  
1a1- .....  
SEQ IDMSLLTLESFESLRVDEAVVILQAGKATQAWHSATGVTV  
SEG .....  
1a1- .....



Prosite for DRF2pites3\_Rml0.2

PS00030	9-117	RNP_1	POCC00030
PS00030	111-119	RNP_1	POCC00030

Pfam for DRF2pites3\_Rml0.2

NAME	RNA recognition motif. (aka RRM, RBD, or RNP domain)
DEFINITION	*IFVGLPMDHTEDLSDIYAFQFPLSIHWDQATGASGCFATVETED *YVSL- ***E LA *PS*FG I-S***M E GRG-CF*IV F *
Query	74 LTVMLDQIDDELAKAFSPFUTTSARVNM--EGGASRGFGVCFSS 120
NAME	KEDEALISDHWGHWGCRIRV*
Query	121 PEGATKAVTDGHWGIVATEPLIV 143

Pedant information for DRF2pites3\_Rml0, frame 3

Report for DRF2pites3\_Rml0.3

(LENGTH)	235
(MW)	26308.08
(PI)	8.95
(HMMOL)	8H[SPROT:PAR1_HUMAN POLYADENYLATE-BINDING PROTEIN 1 (POLY(A) BINDING PROTEIN
(1) (PASP 1):	1e-113
(FUNGAT)	04.05.05 rna processing (5'-end, 3'-end processing and rna degradation) [S.
(FUNGAT)	VER165w] 1e-64
(FUNGAT)	30.03 organization of cytoplasm [S. cerevisiae, YER165w] 1e-64
(FUNGAT)	05.04 translation (initiation, elongation and termination) [S. cerevisiae,
(FUNGAT)	YER165w] 1e-64
(FUNGAT)	30.10 nuclear organization [S. cerevisiae, YER165w] 1e-64
(FUNGAT)	03.18 recombination and dna repair [S. cerevisiae, YER165w] 1e-24
(FUNGAT)	11.04 dna repair (direct repair, base excision repair and nucleotide excision
(FUNGAT)	repair) [S. cerevisiae, YER165w] 1e-24
(FUNGAT)	04.05.99 other rna-transcription activities [S. cerevisiae, YER165w]
(FUNGAT)	2e-19
(FUNGAT)	04.05.03 rna processing (splicing) [S. cerevisiae, YER165w] 2e-14
(FUNGAT)	04.01.04 rna processing [S. cerevisiae, YER165w] 1e-11
(FUNGAT)	11.01 stress response [S. cerevisiae, YER165w] 1e-11
(FUNGAT)	99 unclassified proteins [S. cerevisiae, YER165w] 1e-09
(FUNGAT)	04.07 rna transport [S. cerevisiae, YER165w] 1e-09
(FUNGAT)	30.13 organization of chromosome structure [S. cerevisiae, YER165w] 8e-09
(FUNGAT)	98 classification not yet clear-cut [S. cerevisiae, YER165w] 2e-09
(FUNGAT)	03.13 meiosis [S. cerevisiae, YER165w] 2e-08
(FUNGAT)	04.99 other transcription activities [S. cerevisiae, YER165w] 3e-08
(FUNGAT)	03.01 cell growth [S. cerevisiae, YER165w] 3e-08
(FUNGAT)	06.04 protein targeting, sorting and translocation [S. cerevisiae, YER165w]
(FUNGAT)	3e-04
(FUNGAT)	08.01 nuclear transport [S. cerevisiae, YER165w] 3e-04
(BLOCKS)	BL000308 Eukaryotic RNA-binding region RNP-1 proteins
(BLOCKS)	BL000900 Bacteriophage-type RNA polymerase family proteins signature
(SCOP)	d1x1_4.34.7.1.3 Sex-lethal protein [(Drosophila melanogaster)] 9e-23
(SCOP)	d7u1_4.34.7.1.2 D1A protein (human (Homo sapiens)) 6e-24
(PIRAX)	d1x1_4.34.7.1.1 Nuclear ribonucleoprotein A1, RNP A1, U7 1e-13
(PIRAX)	nucleus 1e-110
(PIRAX)	duplication 1e-110
(PIRAX)	RNA binding 1e-110
(PIRAX)	nucleolus 4e-10
(PIRAX)	random repeat 4e-10
(PIRAX)	single-stranded DNA binding 1e-06
(PIRAX)	DNA binding 9e-12
(PIRAX)	phosphoprotein 4e-10
(PIRAX)	mitochondrion 4e-07
(PIRAX)	heterotrimer 4e-06
(PIRAX)	alternative splicing 1e-15
(PIRAX)	chloroplast 3e-11
(PIRAX)	transcription regulation 3e-09
(PIRAX)	GTP binding 2e-06
(SUPFAM)	helix-destabilizing protein 1e-07
(SUPFAM)	nucleolin 4e-10
(SUPFAM)	glycine-rich RNA-binding protein 2e-07
(SUPFAM)	yeast RNP1 protein 2e-08

[SUPFAM] unassigned ribonucleoprotein repeat-containing proteins 3e-25  
[SUPFAM] polyadenylate-binding protein 1e-112  
[SUPFAM] ribonucleoprotein repeat homology 1e-112  
[PROSITE] RNP\_1 1  
[PFAM] RNA recognition motif. (aka RRM, RBD, or RNP domain)  
[RM] All\_Beta  
[RM] 3D

SEQ EKSRLVCLRAAVFHWSPSTPSYPTASLVGDLHPNTKAMLYETSPAGPILSICROL  
Inel- .....EEETTTTCHHHHHNGGCCKEEESEET  
SEQ ITSGSHYAVVFWHTKDAHALOTWTFVIGKGPVIRHNSORDSLKSGVQNIPTKIL  
Inel- TTTCEEEEEEKCKKHGHHKHHTTEE--TT---EEEEEETTTTCCCKEESEEC  
SEQ DRSIMWALDTVAFQWILSCVYVCDKSGSGYGVFVETKEAENIKKHQKLLGR  
Inel- TTTTCHHHHHNGGCCKEEESEETTTTCEEESECKHHHHHH.....  
SEQ KVPYQPKSRKEAEALGARAKEFPWYTIKPTGDMDEKRLDLPKPGPALSVH  
Inel- .....

Prosite for DMFSpites\_010.3

PS00030 152->160 RNP\_1 P00C00030

Plan for DMFSpites\_010.3

NAME	RNA recognition motif. (aka RRM, RBD, or RNP domain)
RRM	*ITVGLPWDTTEELDIFaQGPVsiIWMIDWATGASGFAVFEED
Query	*TVGL+DTE+L+PS+GPI+SIR+RD+T+S+AV+V+*
RRM	LYVGLWDTTEELTETSPAGPILSICROLITGSHYAVVFWH 75
Query	EEDEAKLIdenQmFWCRIRV*
RRM	DAE A+D+RH ++ G+++R+ 98
Query	76 TDAHALOTWTFVIGKGPVRI
RRM	*ITVGLPWDTTEELDIFaQGPVsiIWMIDWATGASGFAVFEED
Query	I+VGL+ +++ L D S EG I+S+++ D + S+G+V+ PS+ 161
RRM	IFVWLDKSIWALDTVAFQWILSCVYVCD--DKSGYGVFVET 161
Query	EEDEAKLIdenQmFWCRIRV*
RRM	*EAE+RI +HGRH++GR++ V 184
Query	162 KEAENATIDWQKLLNGRAVYV 184

WO 01/12659

PCT/IB00/01496

DKF2phtes3.sp1  
-----

group: testes derived

DKF2phtes3.sp1 encodes a novel 412 amino acid protein without similarity to known proteins.

No informative BLAST results: No predictive prosite, pfam or SCOP motifs.

The new protein can find application in studying the expression profile of testis-specific genes.

unknown

2 EST hits (both from testis libraries)

Sequenced by MediGenomix

Locus: unknown

Insect length: 2899 bp

Poly A stretch at pos. 2870, polyadenylation signal at pos. 2852

```

1 CGAGACGGCC CTGGGGTGCT GGTGGGGTGC CTCTGCTGGG CCTGAGGAAA
51 ACACCTGCCA TGGGCGAAGG CCGGAGGCGC GAGGAGGCC CCCACTGGCG
101 CGGCGGGGCG TCTTGTGCGG TGAGATGGGT GCAAGGATTC CTTAGACGGA
151 ATGTTCTATT CTACACGACA ACACCAATTG CTACCTCTGT GGAATTTATG
201 TAATATTATY TAATATTGAA ACCAGAGAAA AGACTATCTG GCACTTTGAT
251 AATGGAAATG TGGGGCTCAT GGCACCTTAC ATGCTCTCTG AACTTTGGCG
301 TTTTCTTGAC CGGAAGCTAA AACCTCTCAT CTACGTATAC AGCTTTCCAG
351 GATTGACGAG AAGGACCAAA TTGAAAGGCA ACATTCTCTT GCACTTACCT
401 TTACTTTTAT TCAATTACTG TGGCACCTAC CTGGCTAGTT ACTGCTCTCT
451 CCCAGATTTT GAACCTGGCC TTGGGAACCT GGAATCCAGT ATCATTTTGT
501 GTAGAGATTC ACAGCTCTGA ATGAGTGTGA ACCAAATCTC TTTTACCCCG
551 ATGAATCTGG GCGAGCTGTG CTATATCAAT CCAGATACAG TGAGGCTGTG
601 GACATTTGGA AGAATTAACG AGAGAGATTC TTTGACAGCA AGCTGGTGGG
651 AATTACCTCT AGAGATGGG TCATTTTATA ATGAAGAGGA TGTGTTTTTC
701 CCCCACTCTG TGGCGAAGGA TCTCATCTAT GATCCCTGCG TGLACCTCTC
751 AGCAATCTCC GGGCTGGTAG CGAAGAGGCG AGACACTTTC CGCCGCAAGG
801 ATGATCTATA TCTTTGCTTT CAGCCGACCT TGCATTGCTG GACTCCACCA
851 ACTGACTTGT ACATTGGCTG TGACAGAGGT CATCTTTTAA TGATTATGCG
901 AGACACTCTG CAAGTGACTG TACTTAATAA GATAGAGAGG GAATCCCAT
951 TGAAGAGAGC AAGCAATTTT ATCAGCTCAG TACCTCTGTT ATATGAGAG
1001 GAGGGCTGCG TGGCTTCTGG AATTGATGGC TTTGTGTATT CTTTATTAT
1051 TAAAGATGGA AGTTACATGA TCGAGGATTT TCTTGAGATT GAAGACCTTG
1101 TGAACATATG GACATTTTCT CCAATTTTAT CACTTTTCTT GATTCAGACA
1151 GACAGGGGAT CTGTTTATAT CTACACTTFT GGTAGAGAGC CAACTTTAAA
1201 TAAATCTCTA GATCTCTGCG ATGGGAATTT TCAAGCAATT GACCTTTATC
1251 CACCTGGAAC CGAATCTTTC ATGACACTTA GATATTCAGG GGAATTTTGT
1301 GTTTGGTGGC TGAGGATTTG TGLTTTGTGA ACAGATTTT ATCTGAATAC
1351 CTAGACAGCG GTTCTGGCTT GCTTCTCTAT CTCCCTCTCT GCAAGCCGTG
1401 GCAGGAGAGA TGGCTGGCTC TACTTCTATC GGTATATAGA TAAGGAATCC
1451 CTTGAGTCTG TGACAGAGCG CTCTCTCTCT GAATCTCTCG TGCAGACCT
1501 CGTGAAGTCT GTTCTGCTCT CAGAGAGCGC CTCCCTGTCA CACCCTCTCT
1551 TTGAATATTC TATGAGAGCG ATCTCTCTCT TTAATTTTAA CTTTAGGTG
1601 TTTTATTGTT TTTGAATCTT AATATATCCA CACACTTTAA CACTCAAAAG
1651 GTACAGAGGG CTGTTAGTGA AAGTACCCCG CATAGCCAGG TCTGCTCTTG
1701 CAGGAGGCTT GGTACCAATC TCTCATCTCT CTCTTGAGAT GTTTATATCA
1751 TGACACAGCA AAACATAATA AGCACTCTCT TTTACTTGTG TCAATGGCCA
1801 TGATGTGTCT ATAGTGTGCC AGGCACTCTT GCTGTATPAA CTCCATGAGG
1851 TAAACACTCT TGTGCTCTCT ATTTGACAGC TGAGGAGATG AAGGACAAAG
1901 GATTTAAAT AACTTGTCTA ATACTACACA GATATGAAAT GCAATATCTT
1951 GGAATTGAAA CCAAGGTAGT TGGGCTGGAG AGTCACTGCG TTTGCTCTTA
2001 AAAGGAGAAA ACATATGACA ATGCCCTATT TCTTTTCTCA CTTAATCTTA
2051 TATCTTGAGG ATGTTTATAT ATCCACACTT AAGACAGAGC CTAATTTATT
2101 GTATAGCCAC ATAGATATCC ATTATATGAA TATACTATCA TTTTFTAAAA
2151 AGGATATATT ATGACACACT TGAATATTTT CAAACTTTT GAAGCAATAC
2201 TTTTAGATG ATATATAGA GACATTAGAT TTGGACTCTT AGGTGCTATC
2251 ATTATATCTG TTCTTTTTTA ATTTATTATA TTATTAGTGA TTAATAGAAA
2301 CAGAGATTG TATCTGCTT TACACTTGA GATCACTCTA GCTCTGCGCA
2351 TGTATCTCTC AAACACACAG TCAGAAAGCT GTTATCTCTA TCCCTATTAG
2401 ACAAATTAAG GAATTTAGGG TTACAGAGCT GAGGAAAGC ATTTGCGAG
2451 ATTACACTT ACACAGCTAG CACACTGAGG AGCTGGCCCT GGCACCTGGG
2501 ACTGCGCAGC TGCACACGCC TACCTCACTG GGAAGAGATG GATACCTCTC
2551 TTTGATTATC CCGCTGCTCT TCTGACACTT CATTTTGTGT TGCCTTTCTC
2601 TCTCAGATAC CTCTATCTCT AATTACATC TCTCCACTTT TCTTAATTTG
2651 ATAAAGTGTG AGCATGTTT CACTTCAATC TCTCTCAC TGCACGATAC
2701 CAGACAGGG GTAATGAAT GTACACACCA CCACATATTT GAGAAATGCT

```

PCT/IB00/01496

988

WO 01/12659

PCT/IB00/01496

Defigntes1\_9e22

group: testes derived  
Defigntes1\_9e22 encodes a novel 227 amino acid protein with weak partial similarity to Ring-finger proteins.

For the novel protein, Pfam, but not Prosite predicts a C3HC4 type RING finger motif.  
No informative BLAST results; No predictive prosite, pfam or SCOP motifs.  
The new protein can find application in studying the expression profile of testis-specific genes.

similarity to zinc finger proteins

Sequenced by DMFS

Locus: unknown

Insert length: 1318 bp  
Poly A stretch at pos. 1308, no polyadenylation signal found

```
1  CCTCCGCCCG  CTTTCGGAGC  CCGGGGGGGG  CCTGTGGGGG  GGGGAGCCCG
51  GCGCGGACTG  GCGCTCTTTC  GACCTTGAGG  GGAACATGCG  GTTTCCTTCG
101  GATCGTTTGA  AATCTAATGT  TTGGATGCG  GCGCGGGG  CTTGCTCTTC
151  CCGCGCGCGG  GGTTFITTCG  TTTTTCCTTT  TTGCTTTTTC  TCTTTTCTTC
201  CCTCCGGGTC  TCTTTTTCGA  CTCCTTCGCC  CTTTATGCTC  GCGCAGGCTC
251  CCGCTCTCTC  CTGAGAAATG  GGGGAGGCTC  TGGGCTCTCA  GGTTCGCGCC
301  CCACCGGGGC  CCGGGCGAGC  ATGGGGGCGA  AGCAGAGCAC  GCGCGCGCGC
351  TCCCGGAGCC  CTTTCGCGGG  GGTCTCGACG  GATCAGAGCC  CGGTGCGGCG
401  GCGGGAGGGG  GCGGCCCATT  TCGGGCACTA  CCGGACGGGC  GCGCGGGCCA
451  TGGGGTGGG  CAGCGCGTCG  CTCAGCTCGG  TGGCAGGCAT  GCGCATGAGC
501  CCGCAGAGCG  CCGGGGGGCT  GCTCTTTCGG  CTCAGAGCCG  CGGCTCTCGC
551  GCGCAGCGCG  GACTCCGAGA  GGGCGAGCGG  GCGCGAGGGG  TCTGCTTCGG
601  ACTCAGCTTA  TCGCATGGCG  AATGGTTAGC  AGGAGAGGGG  CGAGGTCTAC
651  CATAGAGAGC  GGATGCTGTA  CTTGGGCTCC  CGAGCTTCGC  TGGCGGATGC
701  TCTACTCTCG  CACATCGCAC  CGAGTGGCTT  CAGCTGTCAT  AGTGCTTCTA
751  AGTCCCTCAT  TTGCTCGAAG  TCTCTGCTTC  CTAGAGCAT  GGAATGAGAC
801  TTTATATATG  GTTTCAGCAA  ACTCGGCTTC  TCTTACAGCG  ATGATGCTCT
851  GACTAAGAGC  GCGGTGTAGT  GTTGTATCTG  CTTGAGAGAG  CTTCTCGAGG
901  GGGACACGAT  AGCCAGGCTC  CCGTCTCTCT  GCATCTATCA  CAANAAGCTC
951  ATAGACTCTG  GUTTTGAAGT  GACAGATCTT  TTCTCGAGAG  ACCCTCGGGA
1001  CTGACTCTCG  GCTTCTGTTG  CTGACTCTCT  TCAAGGGGAG  AGAGGCTCCC
1051  TGCTTCAGGG  AGAGGCTCA  CCGGACCTCT  GGGCAGAGCT  GAGCTTGGGA
1101  CACAGGCGGG  AAGAGGCGAC  CCGCTCTTGA  CTGACTCTCA  GATATATGTT
1151  CTGCTCTCTC  CCGTCAGGAG  ACCAATTTGG  ATGAGAGCAA  GTTTGAGAGA
1201  AGAATGATC  ACTGCTATCT  CTTCCTCTCA  CGCTCTAGCT  CAGAGGGGAA
1251  AGGGCATTTT  CTTTTCATC  TTTGAAGGCG  ATTGTGGCTC  TGTCTTAA
1301  GTTTTTCACA  AAAAAAAA
```

BLAST Results

No BLAST result

Medline entries

No Medline entry

Peptide information for frame 3

ORF from 321 bp to 1001 bp; peptide length: 227  
Category: similarity to known protein  
Classification: unclassified

```
1  MGGKQSTAAK  SAGPPFQVST  DGSAYFFPGG  AFHFGHYATC  GGAGMLRSRS
51  VSVVAGMGND  PSTAGGVPTG  LTPPARNCTG  CSERAPGGGG  SASDSTYANG
101  MYGTQTCGQ  WPDQVLYCG  RASLALALPL  HIRPNVPSH  SPTKCVICIS
151  SVASDEKEMH  FIMLEKPKL  SYWQVLTED  AGEDVICLEE  LQAGTIAFL
```

WO 01/12659

PCT/IB00/01496

Z01 PCICIVHSC IDSWEVHRS CPEKPAD

BLASTP hits

No BLASTP hits available

Alert: BLASTP hits for DRF2phea3\_9a22, frame 3

TRZMBL:AF078822.1 product: "RING-H2 finger protein RHA2b"; Arabidopsis thaliana RING-H2 finger protein RHA2b mRNA, complete cds., M = 1, Score = 111, P = 2.8e-06

TRZMBL:AF078822.1 product: "RING-H2 finger protein RHA2a"; Arabidopsis thaliana RING-H2 finger protein RHA2a mRNA, complete cds., M = 1, Score = 112, P = 6.4e-06

TRZMBL:AC004138.14 gene: "T17H13.17"; Arabidopsis thaliana chromosome II BAC T17H13 genomic sequence, complete sequence., M = 2, Score = 123, P = 1.4e-05

PIR:T02286 hypothetical protein T1308.23 - Arabidopsis thaliana, M = 1, Score = 142, P = 8.8e-08

>PIR:T02286 hypothetical protein T1308.23 - Arabidopsis thaliana  
Length = 327

NSFs:

Score = 142 (21.3 bits), Expect = 8.8e-08, P = 8.8e-08  
Identities = 24/57 (42%), Positives = 30/57 (52%)

Query: 166 SPRLSTYNDVLTQDAGECVICLLELLQDQIARLPLCLIVHSCIDSWEVHRSQ 222  
S P + + L T D + C + C I E E + G L A C 1708 C I M + M S C P  
Sbjct: 206 SLPSVKITPQHLTKHNSQCTVCHSEFIYGGDTELPCHIVHSCIVVPLALHNSCP 262

-----  
Pident information for DRF2phea3\_9a22, frame 3

Report for DRF2phea3\_9a22.3

(LENGTH) 227  
(MW) 23782.62  
(PI) 6.18  
(HOMOL) PIR:T02286 hypothetical protein T1308.23 - Arabidopsis thaliana Zs-08  
(FUNCTION) 19 unclassified proteins [S. cerevisiae, YDR313c] 4e-06  
(FUNCTION) 30.07 organization of endoplasmic reticulum [S. cerevisiae, YOL013c]  
(FUNCTION) 0.001  
(FUNCTION) 06.13 proteolysis [S. cerevisiae, YOL013c] 0.001  
(PFAM) Zinc finger, C3HC4 type (RING finger)  
(NM) Irregular  
  
SEQ HGGQSTAAARNGPFGVSTQELAVPPGGAPHTGRTYTGGAAGLASNSVSVAGQHGQ  
PRD ccc  
  
SEQ PSTAGGVFFGLTPASNGTGDRAPOGGGASDSTYAMGVOYTGQGHRRDQHLVLS  
PRD ccc  
  
SEQ RASLADALPLHTAPRWFSHSGFPCFCSASVSDENDQFICLSFRLSYNDVLTQD  
PRD HNNNNNNHNN  
  
SEQ AGGVICLLELLQDQIARLPLCLIVHSCIDSWEVHRSQPEKPAD  
PRD ccc

(No Prosite data available for DRF2phea3\_9a22.3)

Pfam for DRF2phea3\_9a22.3

NAME Zinc finger, C3HC4 type (RING finger)  
HMM "CPICFCTTQIDyWFFdeWHLICkspPcypCirrW.....Cmc"  
C IC L+++ D++ LPC+ ++ +CI +M CP+  
Query 184 CVIC-----LEELLQDQIARLPLCLIVHSCIDSWEVHRSQPEK 224

WO 01/12659

PCT/IB00/01496

DKF583hes3\_9120

group: testes derived

DKF583hes3\_9120 encodes a novel 205 amino acid protein with similarity to human KIAA0336 gene.

No informative BLAST results: No predictive prosites, pfam or SCOP motifs.

The new protein can find application in studying the expression profile of testis-specific genes.

unknown

complete cDNA, complete cds, EST hits

Sequenced by DKF5

Locus: /map="44.1 cR from top of Chr17 linkage group"

Insert length: 2509 bp

Poly A stretch at pos. 2499, polyadenylation signal at pos. 2481

```

1 CTGCGCGAGA TGACCTGGGC ACCTCTGGCT TGAATCGGGA AATACTGATC
51 AAGCGGCATT TATTCTCTCT TCAAGAACTC TAAGCTAGAC AGAGAAGATG
101 AGGCGGAGGA AGTTCATCAA TGCTCTGGCT GGAGGACAGC CAATTCGAG
151 ACATTGGCAA CGGAGTGATC AAATCATAG ATCATGAGGC CTAAATGAA
201 TANGGAAGAA AGAAGATGCG CAGAGGCTGA GAACAGAAAG AGAGGCTGGA
251 GGGGCTCTAA ATCTTGAAIA TTAGGTTATA ATATACGTCT ATGGGTAAAG
301 ATTGGAAGAA TTGTGTAGGA GCGAGTAGTC AAAAAGTAGA AGCAGTTTGG
351 AGAGTAGATT ACAATATACA AGAGGCAAGT GCGTAAAGG TGGAGCTATA
401 GGTCAATTGA GCTCAAGAAA CTGAGTCTCT AGGCGATTGG TTAGTCTATC
451 TGTCTAGACT TCAAGTTTGT CTAGAGTGA AATTACAGAG ACTGATCTGT
501 GCGAAGCTCA CAGTTTCTTC AGCACTCAAA ACACATAGAG AAATATAGCC
551 AAGATGTCTG TGGATCAAT GACCTACGAG GCCCGTTTCT TTGCTTCAAC
601 GCGCAAGAGC TGCATGCTTC GATCTCTACT TCGATTTTCA GACTACCTAT
651 TTGAAGTAGT CGAGGCGCTT GACACAGTTA TTCTGAAGAA GCTGAGTGGC
701 ATGCGAAGCT GTGACATTAG CCGAGTGGAG ATTGCGAAT GCAACAGAAA
751 GTTCTTTTTC TTGATGAAG GACATTTTGA TAACTTTTTT AGCAAAATGG
801 AGCAACTGTT TTGCAAGCTG ATTTTACGTA TTGCTCAAAA CATCTCTGTT
851 CTTGAGATA AATCTAAGGA GACACTTTAT AGTGAAGAGG ATTTCAGCA
901 TTCTCAGAAA GAAATTGAAC AATTACAGGA GAATACAGAG ACTGATTTAT
951 CTACTTAGCA GGGGCTTCTT CGAGATTAG AGAGGAGAAA AATTTCTCAG
1001 GCGAAGCTCA ACGAGAGCTT GACTTCTCTT GATGAGCTTC ATATCTTTGG
1051 CAGAGATCAT GGCAGTAGTG ATTTTAGGGA GAGTTTATGA TCCCTGCTTC
1101 AGAATCTCAG AAATCTCAG AACTATTACG ACATCTGTGA AAGGATGCTG
1151 AAGCGACTGA AATATCTTTA ATTGCTCAGT AGTCAAAAGG AGGAGCCTGT
1201 CAABAAGTAG ATCATAGAGG ACTGTTCAAA CCAATAGAGC TGTCAAACTC
1251 ATACGAGTGA CTCTCAAC CAACCATACT TTTTATTAGA TTCTCTTTGT
1301 CACTCTTTTC TTGATTTCTG TTTTCTCTCT TTTTGTGTC CACTTTGTGT
1351 AGGATAGAG TTACTACTTT TGAATAGG TGAAGCATCT GAGTCTCTTA
1401 ATAGTGGGA AGGATCAAA CAAGAGAGCC ATGACCATTT AAGATATTTT
1451 GCGAGGTATC ACCTTGCTGA TAACTCTT CTGACCTTGA TTATTTTGGC
1501 TTACTTTTTC GATGAGGACA GACAGAGAAA GGATTAAGCT GGTGGCTCTC
1551 TTATATATAT TATTATTTGT TTGAGAGACA GGTCTCTCTT TGTACGCGAG
1601 GTTAGAGTAG ATTTCAGTGG CACAACTCTG GCTCACTGCA ACCTCTGTGT
1651 CTTAGGCTCA ATGATCTCTC CTGCTCTAGC CTGCGAATA GCTAGAGACA
1701 CAGTCTCTCT TCACTACTCT TGGTAAATTT TTTCAGAGAA ACAGAGCTCT
1751 ACTATATTTT CAGGCTGAG TGGCTCTTTT ATTACACAGT CATTACACTG
1801 GCGAGAGGAG AAGATAGAGT ACTGCTCTCT GCTCTGTGAA TTCTTTTCA
1851 TGAGGAGGTC AATATGAGT GCAAGAGAGC ATGTAGAGAA AAGAGACACC
1901 TTGATCTTAA ATAAAGAGAA AGTTGTTTAA TTGCAAAAT AAATCCCTGT
1951 ACAAAGAAC TGTGATCTTT AAGCAATTGA CTCTCTTAGA GTCCAGCAGA
2001 AGACCTTAGA CAAAAGAGC AGACCTCACT GAGTATAGAA AGGAGCATG
2051 TAGATATAG TCACTGTGGA AATTATATT TACTACTGT TAGATATCTA
2101 TTGCAATTGT TTTCATACT TCACTGTGTT TTGAAATCTG CTTATACCTC
2151 AATATATATT TATTCACACT CATAGCATC AATATATTTA TGGCTCTGAG
2201 GCGAAATTTG TGTTAAACT CAATGGAATC TAATATTCTT TTATCTCTTT
2251 ACTGCTCTGA AATCTTTAGG TCACTGAGG AAGGGGAGA AATAGCATG
2301 CTCTCTCTGA AGTATCTCT TCCCTCTCAT GTTCTCTTAA AGAGGAGAAC
2351 TTGAGGTTTC TCTTTATGT AGAGAGAGAG TAATCTAGGG TGATTTTGCA
2401 ATGAATATTT CATATATATT GAAGCTCTCT CTTATAGTGA AATATCTTTA
2451 TTATCAAGAA GTCTTTTTTC CAATCTCTGA CATTAAATAT ATGTCTTTTA
2501 AAAAAAAAA

```

BLAST Results

WO 01/02659

PCT/IB00/01496

Entry AC004148 from database ENBL:  
Homo sapiens chromosome 17, clone HC3524C3, complete sequence.  
Score = 5245, E = 0.0e+00, Identities = 1049/1049  
3 exons

Entry H5356361 from database ENBL:  
Human STS TIGR-A003929.  
Score = 1005, E = 1.3e-18, Identities = 201/201

Entry H50043 from database ENBL:  
human STS SNGC-74031.  
Score = 935, E = 2.8e-37, Identities = 205/215

Medline entries

No Medline entry

Peptide information for frame 2

ORF from 554 bp to 1168 bp; peptide length: 205  
Category: putative protein  
Classification: no clue

1 MSVDPHTVEA QFEGFTPTQC HLAIITAFDD VLFTVMDAVE QVILKRLQGI  
51 PKKDIPVQI RACTESFLCY HSHVPHLTS HSDLFQGLI LAIPSRILLP  
101 EDKCKETPTB EIDFQRLQRE IEDLOKXTYT ELCTQALLA ELEKQIVQA  
151 ELKQVITFTD ELHVGKONG TSDPFEELVS LVQKRLQQLQ TRHWKESR  
201 ELRLS

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DRFpites3\_9120, frame 2

TREMBL|H5356361.1 gene: "K1A0336"; Human mRNA for K1A0336 gene.  
complete cds., E = 1, Score = 107, E = 0.0081

>TREMBL|H5356361.1 gene: "K1A0336"; Human mRNA for K1A0336 gene.  
complete cds.  
Length = 1,583

HSPs:

Score = 107 (14.1 bits), Expect = 8.2e-03, E = 6.1e-03  
Identities = 42/140 (30%), Positives = 76/140 (54%)

Query: 65 KEFLCPNIGHTWLFPMEDLFLQLILRIPSHILLPDKCRTPYSEED----PQRLQRE 120  
EK CF+K H +NL +D +L R ILL +D +P +D + L++  
Sbjct: 796 EKECTPIEDH-GHSLPLQD--LIDRRAVELIL-NQRLASPTVQKPLASVLEKEE 851  
Query: 121 IEDLOQ--HYTELETRQALLAELEKQIVQALQCTLYFTCLHWVQKCTSDPFEEL 178  
IE L++ K E E K L-A ++ +E + + R-Y T +EL ++ + S+  
Sbjct: 852 IELKAKCKEKEKEIKVILVA-VAAKELDSKSTETVVEELKSLKSE--QQLSASM 908  
Query: 179 VELVQSRLEQIRHWVKESESLFI 204  
L+D + +H+ EE+G+L +  
Sbjct: 909 ADLIQGAESYKMLLEYKQSQQLQV 934

Pendant information for DRFpites3\_9120, frame 2

Report for DRFpites3\_9120.2

[LENGTH] 205  
[MW] 24140.13  
[pI] 5.51  
[EW] All Alpha  
[KW] COL120\_CoIL 18.05 %



PCT/IB00/01496

(No Prosite data available for DKFZp003\_9120.2)  
(No Pfam data available for DKFZp003\_9120.2)

WO 01/2659

PCT/IB00/01496

ORFSpntes3\_9122  
-----

group: testes derived

ORFSpntes3\_9122 encodes a novel 304 amino acid protein with partial similarity to X. laevis  
satanin p80.

No informative BLAST results: No predictive prosites, pfam or SCOP motifs.

The new protein can find application in studying the expression profile of testis-specific  
genes.

similarity to C-terminus of satenin p80

Sequenced by ORF3

Locus: unknown

Insert length: 2676 bp  
Poly A stretch at pos. 2665, no polyadenylation signal found

```
1 CTCTCTAGGC TGGCGGGGCG TGGTCTCAG GGGCGAGGCT GGGCTGAGGC
31 GGGCGGGTAC CATGAGGCGC GGGTACTTAA GAGATTATGC CATCAGAAAC
101 GCACAACTTT AAAAAGCGGA ACTTTCTTAA TAGATTATGC CATCATTCA
151 TTGATCTTCC TAGAAAAAGG ATCTCTAATT TCACTATAAA GAACATGAG
201 GAGCTTTAGA ATCTCTGAAA ACCTTTGGCT GCTTCTATAA ATAGAACAT
251 TGGCAAAACT CTGAAAAGGC CAGATAAACT TGTAAAGTGT ATCTATGCGA
301 GAAGAAAGAT TGTATCTGCG TTCTCAATCT GTTGTATGAG AAAAAGACAG
351 TGGCTGTGAA GTGGGGGCTG TGACATGGCA KATAAGAAA ATGACATGCG
401 TTGTGAGGCG CAGCTGCTCG AAAAATGCA CAGATGATGT GGAAGATAT
451 TGGTTAAGTC CAGTGTATCT GGTCTTCCAC AGACAGAAAG CCAATCATCA
501 AAATATATGT GCTTTTCTTC TGAGTTTCT CAGGACATG AAACAATGCG
551 CCAAGTTTTC TCGAGAGGGA ATATGAGATT GAACTGATCT TTACTTTCT
601 GAGAGAAAGG AAGTATAAGT GAACCTGTAG CTATTCTTGT GAGGATAGAA
651 GATCTTGGCG TTGTGTTGCA TTGCTTCTCT GTGCTGACCA ATTTCTTACA
701 GGAAGAAAAA CAAATATATCT CAGCTTGGCT GTGTGTGAG TTGTTGCTCT
751 TACTTAAGTC ACTATCTTAA AGCAAAATTC AGAATTTCT TATGTGTGCT
801 TTAAGCTGCG TTCAAGCACT CATTAAGAG TGGTGTGAG AACATCATC
851 CAAGACAGA ATTAATAATG ATGGAATAT TCAATTTTA AACAACAT
901 TAACTGATTT ATGGGAACAG GAAACCATC TTACTTTGCT TCGAGATAT
951 ACTGCTATA TAGCTAAGGA TGTAGATCT TATTTATAC AGTATGATG
1001 AGAGATCTCA TCTACTAAGG AGCAATGCT TTTCGAACAG ATCTCTGAG
1051 TGTATATATT ACAAAGAAAA AAGTCTGCT TGAGAACTGT GAACCTGTGA
1101 AGAATTCGAA ACTATTTTCT TTCTTAAGAA GCGAGATATY GAAGCACTA
1151 ATGAATTCGC AGCAATCTGC TTCAATCTCA AGTGAAGAAA TATCGAAGAG
1201 GAGAGCTTCC AATTCTATTC AGTGAAGAT GCACTATGCA GATTTCTAC
1251 CTTTGTCTGA TTGCGGAGTT ATATGTTAT TTGTACATCT TAAAGACTAC
1301 TGGATTTTAA TGCATCTCT CATTAAGATA TATTTATAC TATCTGAAA
1351 AATTAAGACG GACTTCTCAC TAGAAGCAC CAGAGAAATY CATCTATAC
1401 TTTTATTAGA TTCTGTTTTA TAAAGAAAA AAGACATTA AATGTGTGCA
1451 GCTATTCTCT TATCTGTGAA AGACTGAGG TTAAAGATAT GAAGAAATY
1501 AATATTAAAC ATTTTGTGTT CACATGAGA TACTGTATAT GTAAATTCGC
1551 TTAATTATTA ATAGGCTAT GTTGTATGAT AGCAATATCT GTTAAAGAA
1601 ACTAAACACA ACATGCTCTC TGGCATGATA AATCATGGA ATTAATATCG
1651 GGGTTTACAT TCTGTAGAG TTCTGTGGA ACATCTCTC CACATTTCT
1701 AAACTTGCAG AATATCTTTA GTATCTCTCA TATTTTCTG CACATCTCTC
1751 ATCTCATGTA TGATGTGTT ATGCTATCTT AAGGAAAGG GTGAATATGT
1801 TTTCTATGTA ATGTTTACTY GGAATCTCT ATGAGCTTGG CTAAATTTAT
1851 TTACTTTTAT ATTTGACATA GATTTCTAAT ATTTCTATY CTTGTATCAT
1901 TTAAGCTTCC TTCAATTCAG TAAATCTCAC AATATTTCT ATTTTGTGCG
1951 TTTTATAAAT TCTGATTTTA TATGAATCTT AATCTTTTCT CACTATATAT
2001 GTTTAATAGA GTTACTATCA GTTATTTAGA ATGTTTATCA GTTAATCTCT
2051 ATCTGTATY TTAATATCCA ACATCTTGTG TCACTATCTC CTCTAATGCT
2101 TATATATATA TCAATACAGA CTTATATGCG TCTTTTCTTA AATATGACT
2151 TTTATGCTCA GTGAAACAAA TTCTGATTA TTATACAGCG TCTAATCTCT
2201 ACTTCTCTTC AAAATGATG GGTFTTATA GCACTGTGCG TGGTTCTCAC
2251 ATCATCTCAT CTTATATATG CTGAGAGAG ATTCTGAGAA AGTFTTATG
2301 TAATTGAATY TCTCTGAGAT AGCTCTTCAA GCATTTGAAT GTAAACCTTT
2351 AGAATTTATY GCTTTATGCA CTACTATAG GAATCTGAG CAAATTTCT
2401 GCTCTTAAAT GTTATCTTTC ACTGAGTCTY GGTFTTGTCT AGCTATATY
2451 TATATATATA GATATCTCTC ACATGAGCA TGAATTTATA TAAATGTTA
2501 TTCTCTTAAA TCTTTAGATY ATAAATATY CAAATATG CACTCTGTG
2551 ATTTGAGAG TGAATATY ATAGAGGAGG TTTTATGAG ATGGAATY
2601 GAATCTTAA CTTTATCTCT CTCTGAGAT TTTATGATGA TAAATATTA
2651 CTTTCTCTTG TCTTAAAAA AAAAAA
```





WO 01/12659

PCT/IB00/01496

BLAST Results

Entry H9541356 from database ENBL:  
human 872 nt:-11840.  
Score = 1267, E = 7.1e-50, Identities = 271/281

Medline entries

78227670:  
Katanin, a microtubule-severing protein, is a novel AAA ATPase  
that targets to the centrosome using a WD40-containing subunit.

Peptide information for frame 3

ORF from 87 bp to 998 bp; peptide length: 304  
Category: similarity to known protein  
Classification: unclassified

1 HARETHAVKRWKCHNRIDED FIDLPKXIS APTWERNKEV RESPEQLAAV  
51 IHTYKQCYE SROKLPKVIIT KAKLVNHPFF HCTKRAQSPS GAGCCDAAE  
101 EHELACAGHL PELLAKHDAFT IYVSSDQES KQTERPRAV KQTFREYQD  
151 HETNAQVLS RHNSLPVATY PWRKESISEL VAILAIEDL QVVDCLPMH  
201 YWLGQEKQY IALGCCVOLL PLVKELLESK FEETVVLQW VLQAVIKRW  
251 HELSRTETI KQWIDILAD QLSGLKQEDN HLTVPTTQ WIAFDVDAVL  
301 LQW

BLASTP hits

No BLASTP hits available

Alert BLASTP hits for DFFPhtes3\_9k22, frame 3

TREMBL:AF054021.1 product: "p80 katanin"; *Xenopus laevis* p80 katanin  
mRNA, partial cds.. E = 1, Score = 146, P = 1.2e-07

TREMBL:AF052432.1 product: "katanin p80 subunit"; *Homo sapiens* katanin  
p80 subunit mRNA, complete cds.. E = 1, Score = 150, P = 1.2e-07

TREMBL:AF052433.1 product: "katanin p80 subunit"; *Strongylocentrotus*  
*purpuratus* katanin p80 subunit mRNA, complete cds.. E = 2, Score = 146,  
P = 1.2e-07

>TREMBL:AF052432.1 product: "katanin p80 subunit"; *Homo sapiens* katanin p80  
subunit mRNA, complete cds.  
Length = 655

HSPs:

Score = 150 (22.5 bits), Expect = 1.2e-07, P = 1.2e-07  
Identities = 35/105 (33%), Positives = 55/105 (52%)

Query: 145 SEVSQDHETNAQVLSRHNSLPVATY PWRKESISELVAILAIEDLQVVDCLPMHVL 204  
S++ + H+TM VL SR+ L+ W + L + L D+ VVVD L+ W +  
Sbjct: 489 SQIRKCHOTNCVLTSEKHNLTVAWVWNGDIKTSVDSAVINDLSVVVDLL----WIV SSS

Query: 205 QEKQYISLGGCVOLLPLVKELLESKFEETVVLQWVLQAVIKRW 249  
+ + L C -L+ -- L+SR+E TV D L+ ---R+  
Sbjct: 545 HQRASLKLKGLCTYTPKQTEILAGRTSTVQCTSLKLLIQWF 589

Peptide information for DFFPhtes3\_9k22, frame 3

Report for DFFPhtes3\_9k22.3

[LENGTH] 304  
[MW] 34767.24  
[pI] 9.19  
[EW] All\_Alpha

WO 01/12659

PCT/JP00/01496

```
[SW]          LOW_COMPLEXITY      1.95 %

SEQ  HASETHVKKMPCWRIEDHFDLPREIISFTWMMKDYKSPQLAATIRFTVGVYK
SEG  .....
PRD  ccccccccccccccccccccccccccccccccccccccccccccccccccccccccc

SEQ  SPDKLNVLYRREKVVHFTFNPCTYRQSPGSGGCHANKEMELACAGLPKELIHDSRT
SEG  .....
PRD  GCGNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN

SEQ  YLVHSIDGCSQTESPESATGCTFFEVQDNETHAQVLFERNHMLVALTWKASISEL
SEG  .....
PRD  #####cccccccccccccccccccccccccccccccccccccccccccccccccccccccc

SEQ  VAYLRIEDKGVVQCLPVLTHCQGESQYISLGGCVLLPLVESLEKPFEEYIVGLM
SEG  .....
PRD  hhhhhhhnncccccccccccccccccccccccccccccccccccccccccccccccccccccccc

SEQ  NLQAVIKRWSELESSTETIINQGHQIILKQGLGLWQDCHLTLVPQYTHIARDVYATL
SEG  .....
PRD  hhhhhhhhhnncccccccccccccccccccccccccccccccccccccccccccccccccccccccc

SEQ  LQLH
SEG  ....
PRD  Nccc

(No Prosite data available for DFFphtex3_9422.3)
(No Pfam data available for DFFphtex3_9422.3)
```



Clone ID	Accession	ORF start (bp)	ORF end (bp)	ORF length (aa)	Protein group	Similarity	ORF start (bp)	ORF end (bp)	ORF length (aa)	Localization Predicted	Localization
DFP2544B162	AL136446	4593	5641	2423	453	adipocyte & communication	unknown	4	no predict*	no predict*	Cytoplasm
DFP2544B122	AL136447	1842	70	1336	488	unknown	unknown		no predict*	no predict*	Cytoplasm
DFP2544B1671	AL136447	1484	78	122	82	unknown	unknown		no predict*	no predict*	Endoplasmic reticulum
DFP2544B162	AL136421	1914	246	4631	442	acidic acid management	similar to: Rinc finger protein	1912	no predict*	no predict*	other/unknown
DFP2544B162	AL136421	1208	191	377	129	transport and traffic	similar to: myp binding protein		no predict*	no predict*	Cytoplasm + nucleus
DFP2544B212	AL136423	1915	218	1348	379	protein management	similar to: protein management	23212.1	no predict*	no predict*	Endoplasmic reticulum
DFP2544B213	AL136413	3300	121	699	193	signaling & communication	similar to: Neurotrophin receptor binding domain. Homology with recoverin indicates dependence on calcium dependent activation of guanylate cyclase.	331.6 as from top of Chst linkage group	no predict*	no predict*	Cytoplasm + nucleus
DFP2544B246	AL136444	2084	73	1878	134	metabolism	similar to: GTPase-5, decoupling protein	5	no predict*	no predict*	Nucleus
DFP2544C0162	AL136446	1731	68	1142	381	acidic acid management	similar to: acid binding protein		no predict*	no predict*	other/unknown
DFP2544C0469	AL136446	899	86	898	270	unknown	unknown		no predict*	no predict*	Cytoplasm + nucleus
DFP2544C1362	AL136447	837	137	873	179	metabolism	similar to: molecular clock protein	1612.3-211.1	no predict*	no predict*	other/unknown



Accession	Contig (bp)	Offset (bp)	Offset (bp)	DEP1 (bp)	DEP2 (bp)	Protein group	Similarity	Chromosomal site	Localization predicted	Localization
AF136597	3928	240	397	386		neurones & myotubes	shared the features of the two proteins therefore should be considered to be involved in the cytoskeleton binding to membrane proteins		"Cytoskeleton in plasma membrane"	Nucleus
AF136597	2105	185	425	157		neurones & myotubes	unknown	86.2 kb from 100kb Chr1 linkage group	"no predict"	cytoplasmic
AF136597	1866	180	1640	287		unknown	unknown	745 A, 21; 847 C, 21	"no predict"	Cytosol
AF136597	2835	272	3177	332		unknown	unknown	"no predict"	"no predict"	Cytoplasm
AF136597	1709	105	1468	448		neurones & myotubes	similar to: DRD-100	87.50 kb from 100kb Chr1 linkage group	"nucleus / nuclear envelope"	Nucleus
AF136597	2316	346	546	200		neurones & myotubes	similar to: neuronal calcium sensor		"no predict"	Nucleus
AF136597	2335	29	1449	407		neurones & myotubes	similar to: GTP-binding protein		"no predict"	Cytosol
AF136597	1787	18	941	359		unknown	unknown		"no predict"	Cytosol
AF136597	2005	104	1000	399		unknown	unknown	(4q)	"no predict"	Cytosol / Nucleus
AF136597	2923	343	1381	473		neurones & myotubes	similar to: calmodulin-related protein	200.5 kb from top of Chr1 linkage group	"nucleus"	Cytosol / Nucleus

[illegible]

CloneID	Access	Contig (bp)	ORFstart (bp)	ORFstop (bp)	ORFsize (aa)	ProteinGroup	Similarity	ChromLocation	Localization Predicted	Localization
DFP2646093	AL136641	1027	37	570	178	protein management	similar to: yeast, ADP and NAD <sup>+</sup> are required for import of an N-terminal protein acetyltransferase 1.	20	"no predict"	Cytosol + Nucleus
DFP2646182	AL136612	2444	559	1215	229	unknown	unknown	4922.1-22	"no predict"	Cytosol + Nucleus
DFP2646183	AL136613	987	93	632	148	unknown	unknown		"no predict"	Mitochondrion
DFP2646183	AL136699	1734	159	1133	325	membrane protein	unknown	11614	"no predict"	Nucleus
DFP26461232	AL136686	2292	276	1829	520	membrane protein	unknown	11612 from 11610-11621	"no predict"	Cytosol + Nucleus
DFP26461562	AL136649	2814	75	971	299	structure & stability	similar to: cell cell interaction protein	1	"plasma membrane"	Plasma membrane + Nucleus + Cytoplasm
DFP26461032	AL136615	1467	126	1084	313	signaling & communication	similar to: protein activator of the nuclear envelope protein kinase		"cytosol or nucleus"	Cytosol
DFP26461242	AL136607	4748	511	1194	228	signaling & communication	unknown		"no predict"	Golgi + Plasma membrane
DFP26461216	AL136690	1548	81	635	145	membrane protein	unknown	873.3-875.1 of from top linkage group	"no predict"	Endoplasmic Reticulum
DFP26461792	AL136699	1741	348	610	81	signaling & communication	similar to: phospholipase protein.	11023	"secret pathway"	Golgi + Plasma

Clone ID	Accession	ORF start (bp)	ORF end (bp)	ORF size (bp)	ORF start (aa)	ORF end (aa)	ORF size (aa)	Similarity	Chemical modification	Localization Predicted	Localization
DFPp441206	AL336465	1122	14	921	286			unknown	177.5 kb from top of Chr11 (large gene)		mitochondrion
DFPp441207	AL336466	1713	98	842	275			unknown	177.5 kb from top of Chr11 (large gene)		mitochondrion
DFPp441208	AL336467	1860	10	1650	547			unknown	177.5 kb from top of Chr11 (large gene)		mitochondrion
DFPp441209	AL336468	1409	9	1021	270			unknown	177.5 kb from top of Chr11 (large gene)		mitochondrion
DFPp441210	AL336469	2868	32	1619	496			unknown	177.5 kb from top of Chr11 (large gene)		mitochondrion
DFPp441211	AL336470	600	109	648	180			unknown	177.5 kb from top of Chr11 (large gene)		mitochondrion
DFPp441212	AL336471	1858	154	1440	429			unknown	177.5 kb from top of Chr11 (large gene)		mitochondrion
DFPp441213	AL336472	2775	779	2392	518			unknown	177.5 kb from top of Chr11 (large gene)		mitochondrion

Clone ID	Access	Contig (bp)	ORF start (bp)	ORF end (bp)	ORF size (aa)	Protein Group	Similarity	Chromosomal map	Localization Predicted	Localization
DFP254420822	ALJ34410	2789	10	535	172	unknown	unknown	7	no predict*	unknown
DFP254402316	ALJ34413	1378	337	1118	384	signaling & communication	similar to: GTP-binding regulatory protein	1	unknown*	cytoplasmic membrane
DFP254403132	ALJ34417	1371	187	1015	303	unknown	unknown	922.1-21.3	no predict*	cytoplasmic/unknown
DFP254403344	ALJ34419	1560	207	844	234	unknown	unknown	17	no predict*	endoplasmic reticulum
DFP254402316	ALJ34402	2048	332	1155	384	unknown	unknown		no predict*	mitochondria
DFP254402316	ALJ34443	2378	279	2345	549	protein management	unknown, P18 protein, similar to: ubiquitin	9	cytosol*	cytosol + nucleus
DFP254403216	ALJ34403	2042	73	873	267	unknown protein	unknown		no predict*	cytoplasmic membrane
DFP254403423	ALJ34417	2416	29	1072	348	transport and cyclic	unknown, a lectin character is predicted	3	no predict*	endoplasmic reticulum
DFP25440682	ALJ34071	952	227	849	231	unknown	unknown, contains serine protease active		no predict*	cytosol + nucleus
DFP254404112	ALJ34070	2406	14	345	114	signaling & communication	unknown	107.7-2.1 from M13 Cl87 linkage group	no predict*	cytoplasmic
DFP254404173	ALJ34444	636	36	402	125	unknown	similar to: juncal protein		no predict*	cytosol + nucleus
DFP254404163	ALJ34422	1192	128	1027	301	signaling & communication	similar to: phospho- protein, or protein	9	cytosol*	cytosol

Clone ID	Accession	Gene ID	ORF start (bp)	ORF end (bp)	ORF length (aa)	Protein description	Similarity	ORF start (bp)	ORF end (bp)	ORF length (aa)	Protein description	Localization
DFP5640182	AL136217		2797	302	110	unknown	unknown					
DFP5640183	AL136219		2201	86	134	unknown	unknown					
DFP5640184	AL136244		1446	75	144	cell cycle	similar to SMI gene					
DFP5640185	AL136618		2936	372	392	signaling & communication	unknown, contains a WW domain with binds to a particular protein. Similar to SMI gene. Y and the remainder of the domain is frequently associated with other domains in the same protein in signal transduction processes					
DFP5640186	AL136619		2515	386	139	structure & function	similar to ankyrin					
DFP5640187	AL136619		1736	24	110	unknown	unknown					
DFP5640188	AL136622		1985	234	213	unknown	unknown					
DFP5640189	AL136622		1260	56	132	signaling & communication	similar to the density lipoprotein (LDL) receptors are the major					

[illegible]





Clone ID	Accession	Length (bp)	Insert (bp)	Protein (aa)	Protein group	Similarity	Commentary	Localization
DEP2/1602019	AL30283	1782	766	1316	171	unknown		
DEP2/1602018	AL304931	1076	65	596	186	similar to: mitochondrial protein		
DEP2/1602017	AL304927	1886	23	1865	611	similar to: transport and protein		
DEP2/1701151	AL309215	1897	149	1701	694	similar to: transmembrane protein		
DEP2/1706111	AL317479	22715	79	1899	633	unknown		
DEP2/1706111	AL317480	26218	86	1861	642	unknown		
DEP2/1602011	AL304846	6892	187	1613	1169	similar to: GTPase		
DEP2/1602012	AL304849	6117	187	1618	144	similar to: GTPase		
DEP2/1602013	AL304849	6118	139	1894	1262	cell cycle		
DEP2/1602014	AL304851	1328	179	1149	642	cell cycle		
DEP2/1602015	AL304852	4293	112	1421	770	signaling & communication		

### Table of cDNA clones and related data

[illegible]

## Group cell structure and motility

[illegible]

test_101	are machine characteristics the same for the given topology?	There is believed to function in self organization and for candidates for GAT	Differentiation	over/under
test_102	are the positive probabilities identical for the 2 dimensions		Differentiation	over/under
test_103	are the probabilities to determine new elements		Differentiation	over/under
test_104	are the probabilities to move left and right 0-1-0	moving to water for chondro-osteogenic differentiation	Differentiation	over/under

[illegible]

Group summary conditions derived

Condition	Condition	Condition	Condition
Condition 1	Condition 2	Condition 3	Condition 4
Condition 5	Condition 6	Condition 7	Condition 8
Condition 9	Condition 10	Condition 11	Condition 12
Condition 13	Condition 14	Condition 15	Condition 16
Condition 17	Condition 18	Condition 19	Condition 20
Condition 21	Condition 22	Condition 23	Condition 24
Condition 25	Condition 26	Condition 27	Condition 28
Condition 29	Condition 30	Condition 31	Condition 32
Condition 33	Condition 34	Condition 35	Condition 36
Condition 37	Condition 38	Condition 39	Condition 40
Condition 41	Condition 42	Condition 43	Condition 44
Condition 45	Condition 46	Condition 47	Condition 48
Condition 49	Condition 50	Condition 51	Condition 52
Condition 53	Condition 54	Condition 55	Condition 56
Condition 57	Condition 58	Condition 59	Condition 60
Condition 61	Condition 62	Condition 63	Condition 64
Condition 65	Condition 66	Condition 67	Condition 68
Condition 69	Condition 70	Condition 71	Condition 72
Condition 73	Condition 74	Condition 75	Condition 76
Condition 77	Condition 78	Condition 79	Condition 80
Condition 81	Condition 82	Condition 83	Condition 84
Condition 85	Condition 86	Condition 87	Condition 88
Condition 89	Condition 90	Condition 91	Condition 92
Condition 93	Condition 94	Condition 95	Condition 96
Condition 97	Condition 98	Condition 99	Condition 100







[illegible]

1016



[illegible]

**Group signal) transduction**

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100	101	102	103	104	105	106	107	108	109	110	111	112	113	114	115	116	117	118	119	120	121	122	123	124	125	126	127	128	129	130	131	132	133	134	135	136	137	138	139	140	141	142	143	144	145	146	147	148	149	150	151	152	153	154	155	156	157	158	159	160	161	162	163	164	165	166	167	168	169	170	171	172	173	174	175	176	177	178	179	180	181	182	183	184	185	186	187	188	189	190	191	192	193	194	195	196	197	198	199	200	201	202	203	204	205	206	207	208	209	210	211	212	213	214	215	216	217	218	219	220	221	222	223	224	225	226	227	228	229	230	231	232	233	234	235	236	237	238	239	240	241	242	243	244	245	246	247	248	249	250	251	252	253	254	255	256	257	258	259	260	261	262	263	264	265	266	267	268	269	270	271	272	273	274	275	276	277	278	279	280	281	282	283	284	285	286	287	288	289	290	291	292	293	294	295	296	297	298	299	300	301	302	303	304	305	306	307	308	309	310	311	312	313	314	315	316	317	318	319	320	321	322	323	324	325	326	327	328	329	330	331	332	333	334	335	336	337	338	339	340	341	342	343	344	345	346	347	348	349	350	351	352	353	354	355	356	357	358	359	360	361	362	363	364	365	366	367	368	369	370	371	372	373	374	375	376	377	378	379	380	381	382	383	384	385	386	387	388	389	390	391	392	393	394	395	396	397	398	399	400	401	402	403	404	405	406	407	408	409	410	411	412	413	414	415	416	417	418	419	420	421	422	423	424	425	426	427	428	429	430	431	432	433	434	435	436	437	438	439	440	441	442	443	444	445	446	447	448	449	450	451	452	453	454	455	456	457	458	459	460	461	462	463	464	465	466
---	---	---	---	---	---	---	---	---	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----

[illegible]

1021

[illegible][illegible]



Group were conducted

Test ID	Test Description	Test Results	Test Status
Test 101	Ability to conduct tests	Informational data results; no predictive results; plan or ROP active	Active
Test 102	Informational data results	Informational data results; no predictive results; plan or ROP active	Active
Test 103	Informational data results	Informational data results; no predictive results; plan or ROP active	Active
Test 104	Informational data results	Informational data results; no predictive results; plan or ROP active	Active
Test 105	Informational data results	Informational data results; no predictive results; plan or ROP active	Active
Test 106	Informational data results	Informational data results; no predictive results; plan or ROP active	Active
Test 107	Informational data results	Informational data results; no predictive results; plan or ROP active	Active
Test 108	Informational data results	Informational data results; no predictive results; plan or ROP active	Active
Test 109	Informational data results	Informational data results; no predictive results; plan or ROP active	Active
Test 110	Informational data results	Informational data results; no predictive results; plan or ROP active	Active
Test 111	Informational data results	Informational data results; no predictive results; plan or ROP active	Active
Test 112	Informational data results	Informational data results; no predictive results; plan or ROP active	Active
Test 113	Informational data results	Informational data results; no predictive results; plan or ROP active	Active
Test 114	Informational data results	Informational data results; no predictive results; plan or ROP active	Active
Test 115	Informational data results	Informational data results; no predictive results; plan or ROP active	Active
Test 116	Informational data results	Informational data results; no predictive results; plan or ROP active	Active
Test 117	Informational data results	Informational data results; no predictive results; plan or ROP active	Active
Test 118	Informational data results	Informational data results; no predictive results; plan or ROP active	Active
Test 119	Informational data results	Informational data results; no predictive results; plan or ROP active	Active
Test 120	Informational data results	Informational data results; no predictive results; plan or ROP active	Active

## Prosite Key

NAME: H-phosphatase site.  
CONSENSUS: H-(P)-(S)-(P).  
NAME: Glycerophosphatase attachment site.  
CONSENSUS: S-G-S-G.  
NAME: Tyrosine sulfotransferase site.  
NAME: cAMP- and cGMP-dependent protein kinase phosphorylation site.  
CONSENSUS: (R/S)T-S-(S/T).  
NAME: Protein kinase C phosphorylation site.  
CONSENSUS: (S/T)-S-(R/S).  
NAME: Casein kinase II phosphorylation site.  
CONSENSUS: (S/T)-S-T-(D/E).  
NAME: Tyrosine kinase phosphorylation site.  
CONSENSUS: (R/S)-S-T-(D/E)-S-T-T.  
NAME: H-acyltransferase site.  
CONSENSUS: G-(E/D)E(K/H)Y(Y/W)-S-T-(S/T)A(O/C)-P.  
NAME: Amidation site.  
CONSENSUS: -S-G-(R/S)S-(R/S).  
NAME: Aspartate acid and asparagine hydroxylation site.  
CONSENSUS: C-S-(D/N)-S-S-(P/T)-S-C-S-C.  
NAME: Vitamin E dependent carboxyltransferase domain.  
CONSENSUS: K(L)D-S-S-T-S-E-S-C-S-S-(D/E)S-S-(L/V)M(F/T)-H(H)-(P/Y)W.  
NAME: Phosphopantetheine attachment site.  
CONSENSUS: (D/E)G(S/T)-L(A)R(S/H)-S(L/V)M(F/T)T-A(C)S(Q/H)-S(L/V)M(F/Y)A(G) (D/E)K(S/D)-S-(L/V)M(F/T).  
CONSENSUS: (P/C)T-(S/T)A(C)Q(L/V)M(F)S(L/V)M(A/T)H-(D/E)M(Q/T)A(L/R)I(L)S(L/V)M(W/T)A-(S/L)V(G)T(A/C)S.  
CONSENSUS: S-T-(S/L)V(M/F)A(S).  
NAME: Acyl carrier protein phosphopantetheine domain profile.  
NAME: Peptidylarginine carboxylase lipoprotein lipid attachment site.  
CONSENSUS: (D/E)A(L)S(L)S(L)V(M/F)T-T-A(G/S)-S(L/V)M(F)T-T-A(C)Q(L)A(G/S)-C.  
NAME: Peptidylarginine N-terminal modification site.  
CONSENSUS: (E/R)H(E/Q)T(A/G)-G-(P/T)S(L/V)M(S/T)T-(L/T)S(L/V)M(S)-S(L/V)M(F/W)T-T(A/S)K(I).  
NAME: Phosphoryl group binding site (CAXX box).  
CONSENSUS: C-(D/H)Q-(S/L)V(M/S)-S-S->.  
NAME: Protein splitting signature.  
CONSENSUS: (D/H)R(S)-S(L/V)Y(A)S(L/V)M(Y)S(L/V)A(T)S(H-N)S(T)C(L).  
NAME: Endoplasmic reticulum targeting sequence.  
CONSENSUS: (E/R)H(Q/S)A-(D/E)H(Q)S-S-L->.  
NAME: Mitochondrial C-terminal targeting signal.  
CONSENSUS: (S/T)A(O/C)P-(R/S)S(L/V)M(A/F)Y(S).  
NAME: Green-photosynthetic surface proteins 'anchoring' hexapeptide.  
CONSENSUS: L-P-S-T-G-(S/T)G(A/Y)D(E).  
NAME: Biquitin nuclear targeting sequence.  
NAME: Cyt6 attachment sequence.  
CONSENSUS: S-G-Q.  
NAME: ATP/GTP-binding site motif A (P-loop).  
CONSENSUS: (A/C)S-S-S-G-S-(S/T).  
NAME: Cyclic nucleotide-binding domain signature 1.  
CONSENSUS: (L/V)M(S)Y(K/S)-S-T-G-(D/H)Q(T/A)-S-(G/A)C-S-T-(S/L)V(M/F)Y(S)-S-T-S-G.

NAME: Cyclic nucleotide-binding domain signature 2.  
CONSENSUS: [LVMF]G-E-L(GAS)[LVM]-K(11)-R-[STAQ]-A-L(LVMA)-[STACV].

NAME: cAMP/GMP binding motif.  
NAME: EF hand nucleotide-binding domain.  
CONSENSUS: D-L-[DMS]-[LFTYF]-[DSEHSTQ]-[DHQGHRA]-[QF]-[LVMQ]-[DEHQR]-[DSQ]-K(2).  
CONSENSUS: [DE]-[LVMFYF].

NAME: Apicomplexan actin-binding domain signature 1.  
CONSENSUS: [DQ]-K(2)-[ATV]-[QTY]-K(2)-W-L-M.

NAME: Actin-like domain signature 1.  
CONSENSUS: [LVM]-L(GD)-[LVM]-[DAGRE]-[SAGS]-[DREAG]-[LVM]-[DEAG]-K(6).  
CONSENSUS: [LVM]-L(M)-[SAGS]-[LVM]-[LVM]-W-L(LVM)S.

NAME: Amphiphysin domain signature.  
CONSENSUS: [C]-[S]-C-K(2)-[GAP]-K(7)-R-[GATDSQR]-C-[GATDSQR]-K(3)-R-[GATDSQR]-K(2).  
CONSENSUS: [CE]-K(6)-7-C-C.

NAME: Amphiphysin domain profile.

NAME: Apple domain.  
CONSENSUS: C-K(3)-[LVMFY]-K(3)-[LVMFY]-K(2)-[DEHQR]-[LVMFY]-K(2)-C-K(3)-C-T.  
CONSENSUS: K(6)-C-L(LVMFY)-F-L(FY)-K(11)-K-C-L(LVMFY)-[KAS]-[ST]-K(11)-L(S).  
CONSENSUS: S-Q-L(ST)-[LVMFY]-K(3)-C.

NAME: Rind 4.1 family domain signature 1.  
CONSENSUS: W-[LVP]-K(2)-[SE]-L(LVM)-K(2)-[QR]-K(6)-L(LVMFY)-K(6)-R-[LVMFY].  
CONSENSUS: K(3)-F-Q(Y)-K(2)-[DEHQR].

NAME: Rind 4.1 family domain signature 2.  
CONSENSUS: [H]-V-S-L(SHQRVY)-[SAG]-K(3)-[FY]-[LVM]-K(2)-[ACV]-K(2)-[LAF]-K(2).  
CONSENSUS: [FY]-G-L(SHQRVY)-[LVMFY].

NAME: Rind 4.1 family domain profile.

NAME: C1g domain signature.  
CONSENSUS: F-K(2)-[DQ]-K(4)-[FYF]-L-K(6)-F-K(3)-D-L-T-L-F-L(FY).

NAME: C-terminal cysteine loop signature.  
CONSENSUS: C-C-K(13)-C-K(2)-[SH]-K(13)-C-L-C-K(2)-K(1)-C.

NAME: C-terminal cysteine loop profile.

NAME: CUB domain profile.

NAME: Death domain profile.

NAME: EGF-like domain signature 1.  
CONSENSUS: C-L-K(13)-G-K(2)-C.

NAME: EGF-like domain signature 2.  
CONSENSUS: C-L-C-K(2)-[QF]-[FYF]-K(4)-E-C.

NAME: Calcium-binding EGF-like domain pattern signature.  
CONSENSUS: [DSQF]-L(DSQFQD)-C-K(11)-C-K(3)-7-C-L(DN)-K(4)-[FY]-L-C.

NAME: Laminin-type EGF-like (L2) domain signature.  
CONSENSUS: C-K(12)-C-K(5)-G-K(2)-K(5)-C-K(3)-H-F(Y)-K(4)-L(S)-C.

NAME: Complement factors 5/9 type C domain (FAMC) signature 1.  
CONSENSUS: [GAD]-W-K(7)-L(SHFFY)-[LVP]-L(LVFA)-[HSTDEQ]-K(6)-[LVP]-K(2)-[FY]-L(LVTV)-[KQK]-G.

NAME: Complement factors 5/9 type C domain (FAMC) signature 2.  
CONSENSUS: F-K(8)-K(6)-[LAF]-K-L(GE)-[LVP]-L-G-C.

NAME: Patched-associated (FHA) domain profile.

NAME: Fibronectin type 3 and galectin domain C-terminal domain signature.  
CONSENSUS: W-W-[LVMFY]-K(2)-C-K(2)-[GSA]-K(2)-N-G.

NAME: Type I fibronectin domain.

CONSENSUS: C-u6,13,ELP7)-u73-(FTW)+-(R3)-u8,109-C-u6,9-C.

NAME: Type II *Shewanella* collagen-binding domain.  
CONSENSUS: C-u2,3-P-P-u(FTW)-u73-C-u8,109-W-C-u44-(DNR3)-(FTW)-u1,13-(FTW)+-(FTW)-C.  
CONSENSUS:

NAME: Hemagglutinin domain signature.  
CONSENSUS: (LIFTAT)-u73-W-u2,13-(PE)-u72-(LVMPT)-(DNRQ3)-(TA)-(AV)-(LVMPT).  
CONSENSUS:

NAME: Kringle domain signature.  
CONSENSUS: (PT)-C-S-M-P-(DNR).  
CONSENSUS:

NAME: Kringle domain profile.  
CONSENSUS:

NAME: LDL-receptor class A (LDLR-A) domain signature.  
CONSENSUS: C-V(LMA)-u73-C-(DNR3)-u73-(DNRQNT)-C-u7,4-(ETADE)-(DNR)-u23-u1,13-  
CONSENSUS: C.  
CONSENSUS:

NAME: LDL-receptor class A (LDLR-A) domain profile.  
CONSENSUS:

NAME: C-type lectin domain signature.  
CONSENSUS: C-(LVMPTATQ)-u3,123-(W1)-u-(DNR3)-u73-C-u3,6-(FTW)(LVTAT)-(LVMPTAT)-  
CONSENSUS: C.  
CONSENSUS:

NAME: C-type lectin domain profile.  
CONSENSUS:

NAME: Lact domain signature.  
CONSENSUS: C-u113-A-u2,43-G-u73-C-u73-G-u8,9-P-u73-C.  
CONSENSUS:

NAME: Chondroitin domain signature 1.  
CONSENSUS: C-u-(DNR)-u73-C-u73-G-(R33)-u6,73-P-u6,73-C-u6,73-C-P.  
CONSENSUS:

NAME: Chondroitin domain signature 2.  
CONSENSUS: P-P-u8-(DNR)-u6,73-W-L-u-(DNR).  
CONSENSUS:

NAME: Bomanin domain signature.  
CONSENSUS: C-u6,73-C-u73-C-C-u-(DNR)-(PT)-u73-C.  
CONSENSUS:

NAME: Thymoglobulin type-1 repeat signature.  
CONSENSUS: (FTW)(P)-u-P-u6,73-G-u-(FTW)-u73-Q-C-u6,109-C-(FTW)-C-V-u3,43-  
CONSENSUS: (DQ).  
CONSENSUS:

NAME: P-type "Tubulin" domain signature.  
CONSENSUS: B-u73-C-u-(PTTPT)-u73,44-(PT)-u73-C-u6,9-C-C-(FTW)(P).  
CONSENSUS:

NAME: Cellulose-binding domain, bacterial type.  
CONSENSUS: W-M-(TAG3)-(TDN)-(LVM)-u73-(DRT)-u-(GFT)-u73-(LVMPT)-(DA).  
CONSENSUS:

NAME: Cellulose-binding domain, fungal type.  
CONSENSUS: C-G-G-u6,73-G-u73-C-u73-C-u73,13-(DNRQ)-u-(FTW)-u73-Q-C.  
CONSENSUS:

NAME: Chain recognition or binding domain signature.  
CONSENSUS: C-u6,13-C-u6,73-C-u73-G-u6,73-C-u6,73-(FTW)-C.  
CONSENSUS:

NAME: Barvin domain signature 1.  
CONSENSUS: C-G-(R3)-C-L-u-V-u-M.  
CONSENSUS:

NAME: Barvin domain signature 2.  
CONSENSUS: Y-(DNR)-Y-(DNR)-P-Y-(DNR)-C.  
CONSENSUS:

NAME: BR repeat.  
CONSENSUS: (DEENLYT)-u73-R-u73,73-(FTW)-u1,143-(TAN)-(GLMF)-X-(FTTDA)-X44-  
CONSENSUS: (DEEL)-u73,13-C-u73-C-u6,73-(W)-X-(W)-X-(W)-X-C-u73-(LVM)-  
CONSENSUS:

NAME: WAP-type "low molecular mass" domain signature.  
CONSENSUS: C-u-(DNR)-u73-C-u73-C-C.  
CONSENSUS:

NAME: Planted core / disulfide-linked binding domain.  
CONSENSUS: H-u-(LVMFTW)-u8,113-C-u73-C-u73-(LVMPT)-u8,109-C-u73-C-u6,9-(DNR)-  
CONSENSUS: u73-C-u8,9-C.  
CONSENSUS:

NAME: C2 domain signature.  
CONSENSUS: (ACTQ)-u73-L-u73,13-D-u6,13-(DNR)(LPT)-(TMD)-u-(STAP)-D-(PA)-(PT).  
CONSENSUS:

PCT/IB00/01496

1027

PCT/LB00/01496

1028

1029

PCT/IB04/01496

1030



WO 81/12659

PCT/IB80/01496

CONSENSUS: [LVMK]E[QR]K- $\alpha$ 2-[LVNMP]K-2-[LVN]M.  
NAME: Signal 70 known ECF substrate signature.  
CONSENSUS: [ST]A[V]P[QK]L[L]R[Q]E[LV]L[LVTA]-Q- $\alpha$ -[ST]A[V]-[LVNMP]C-[LVNKAQ]-  
CONSENSUS: [G]ST[A]P-[L]M[P]T[W]Q- $\alpha$ 1,14-[ST]A[P]-[P]Y[V]-[L]P]- $\alpha$ 2-[P]Y.  
NAME: Signal 34 unknown domain ATP binding region A signature.  
CONSENSUS: [L]VM[P]T[D]- $\alpha$ Q-[R]Q[Q]-[T]T[Q]- $\alpha$ -[ST]A[V]-G-S- $\alpha$ 2-[LVNMP]Y.  
NAME: Signal 34 unknown domain ATP binding region B signature.  
CONSENSUS: [K]R-[L]V[M]P]- $\alpha$ 2-[A]-[D]N[Q]A[SH]-[Q]N[Q]K-G-[T]T[M]-[LVNMP]T[D]-[R]E-[E]K-  
CONSENSUS: [L]V[M].  
NAME: Signal 34 unknown domain C-terminal part signature.  
CONSENSUS: [P]Y[V]-P-[Q]Q]-M-[L]V[M]-S-[R]Q]-L- $\alpha$ -[Q]H[A]T].  
NAME: Signal 34 unknown domain profile.  
NAME: Single-strand binding protein family signature 1.  
CONSENSUS: [L]V[M]P-[N]T[T]-[L]R[T]-[L]V[M]- $\alpha$ -[L]V[M]P[D]-G-[N]H[R]K-[L]V[M]-[G]S[T]- $\alpha$ -[G]E[T].  
NAME: Single-strand binding protein family signature 2.  
CONSENSUS: T- $\alpha$ -W-[D]T-[E]H[Q]-[L]P[M]- $\alpha$ -[L]V[M]P-[P]T]-[N]G[X]R].  
NAME: Structural homology DNA-binding protein signature.  
CONSENSUS: [K]R[K]-P- $\alpha$ 2-[L]V[M]P]- $\alpha$ -[E]R[K]Q[A]- $\alpha$ -[T]H[R]T]- $\alpha$ -[G]A]- $\alpha$ -[E]N]-P- $\alpha$ -T.  
NAME: Dna protein family signature 1.  
CONSENSUS: H-[P]Y]- $\alpha$ -[L]V[M]- $\alpha$ -G- $\alpha$ -[T]G-[L]Y]-H- $\alpha$ -[D]E].  
NAME: Dna protein family signature 2.  
CONSENSUS: [L]V[M]P[T]D[M]- $\alpha$ -[L]V[M]-[G]A]-S-R- $\alpha$ 2-[L]P-[K]D[N]- $\alpha$ -[H]-[P]A].  
NAME: DNA repair protein recC family signature.  
CONSENSUS: H-H-S-P-S-Q.  
NAME: recA signature.  
CONSENSUS: A-L-[E]K]-[P]Y-[P]Y]-[T]A]-[T]T[A]D-[L]V[M]Q-S-R.  
NAME: RecP protein signature 1.  
CONSENSUS: P-[E]D]- $\alpha$ 2-[L]V[M]E[D]- $\alpha$ -G-[G]A[D]-P- $\alpha$ 2-S-R- $\alpha$ -[P]Y]-[L]V[M]-Q.  
NAME: RecP protein signature 2.  
CONSENSUS: [L]V[M]P[T]E[D]- $\alpha$ - $\alpha$ 2-[S]-[E]A]-[E]N]-L-D- $\alpha$ 2-[E]M[D]- $\alpha$ -[D]-L.  
NAME: RecB protein signature.  
CONSENSUS: C- $\alpha$ 2-C- $\alpha$ 2-[E]T]- $\alpha$ -[I]-C- $\alpha$ -I-C- $\alpha$ -[E]-R.  
NAME: Human H2A signature.  
CONSENSUS: [A]C]-G-L- $\alpha$ -P-P-Y.  
NAME: Human H2B signature.  
CONSENSUS: [E]D]-P-[L]V[M]-[R]Q]-T- $\alpha$ 2-[E]K]- $\alpha$ -[L]V[M]E[D]- $\alpha$ -[P]A[D]-[D]Q]-L- $\alpha$ -[E]K]-H-A-  
CONSENSUS: [L]V[M]-[T]A]-E-Q.  
NAME: Human H3 signature 1.  
CONSENSUS: E-A-P-E-Q-L.  
NAME: Human H3 signature 2.  
CONSENSUS: P-P- $\alpha$ -[E]A]-L-(Y-A)-[E]P[Q]-[D]K]-[P]Y].  
NAME: Human H4 signature.  
CONSENSUS: G-A-E-E-R.  
NAME: HMG1/2 signature.  
CONSENSUS: [P]R-S-[E]K]-E-C-S-[E]K]-S-W-E-T-M.  
NAME: HMG14 and HMG17 DNA-binding domain (A + T box).  
CONSENSUS: [A]T]- $\alpha$ 1,2]-[R]E[K]-[G]P]-S-G-S-P-[E]K]- $\alpha$ .  
NAME: HMG14 and HMG17 signature.  
CONSENSUS: R-S-S-A-S-L-S-A-[E]K]-P.  
NAME: Bacteriophage signature.

WO 81/12659

PCT/80081496

CONSENSUS: (STAVYV)KTD-R-4(L)DHS-4L-7(L)DEKOT7P-Y-RPT1-KD(LVMAF7)K2-  
CONSENSUS: (LVMA)K14(LVMA)K48-R-Y-K(1,13)(LVMA)K23-N-(KACF)K25(PF1)  
NAME: Chromosome profile  
NAME: Chromosome domain signature  
CONSENSUS: (PT12)K(LVMAK)K(R)W-K(GDWR)K(YVWL2)K(5,6)ST7-W(22)K(PTDR)K21-  
CONSENSUS: (LVMAK)  
NAME: Chromosome and chromosome domain profile  
NAME: Regulator of chromosome condensation (RCC1) signature 1  
CONSENSUS: Q-K-H-D-K2(LAV)K-L-G-R-K-T  
NAME: Regulator of chromosome condensation (RCC1) signature 2  
CONSENSUS: (LVMAF)K21(AKCE)K2-Q-K23-N-(STAG)K3(LVMAF)K4+(LVMA)  
NAME: Pannexin P1 signature  
CONSENSUS: (AVY)K(20PFY)K-K23-3(STT)K-S-K-4  
NAME: Histone transition protein 1 signature  
CONSENSUS: S-E-R-Z-Y-R-E  
NAME: Nuclear transition protein 2 signature 1  
CONSENSUS: H-K23-H-K(DN)K-S-K-P-Q-E  
NAME: Nuclear transition protein 2 signature 2  
CONSENSUS: E-K-R-E-K23-R-G-E-K23-E-(R)K-E  
NAME: Ribosomal protein L1 signature  
CONSENSUS: (R)K23(LVMA)K-K23-3(LVMA)K-G-K23-(LMS)K(DNR)K(PTK)K(RSA)YK-G-K-  
CONSENSUS: (LMP)K(DNDOT)K  
NAME: Ribosomal protein L2 signature  
CONSENSUS: F-K23-R-G-ST(LVY)K23-N-(L)PK-K(DK)  
NAME: Ribosomal protein L3 signature  
CONSENSUS: (FL)K48(LDNR)K23-1A(DS)K-STT-K-Q-(R)K-G-K23-G-K23-R  
NAME: Ribosomal protein L5 signature  
CONSENSUS: (LVMA)K23(LVMA)K(STAC)K(GK)KQYK-K23(LVMA)K-STT-K-STACQ(R)K-  
CONSENSUS: (PTA)  
NAME: Ribosomal protein L6 signature 1  
CONSENSUS: (PS)K(DDR)K-Y-R-(SA)K-K-G(LVMA)  
NAME: Ribosomal protein L6 signature 2  
CONSENSUS: Q-K23(LVMA)K-K23-(R)K-K23-R-K-P-K-G-(LVMA)Y(LVMA)K-K23-(R)K  
NAME: Ribosomal protein L7 signature  
CONSENSUS: G-K23(DNR)K43-Y-K23-G-(PT)K-K23-N-(PT)K-L-K3-(DA)K-K23-STTN  
NAME: Ribosomal protein L9 signature  
CONSENSUS: (DSD)K-K23-(DS)K(LVMA)K-STTN(YV)K-A-K-(DSD)K(LVMA)K-K23-(R)K-R  
NAME: Ribosomal protein L11 signature  
CONSENSUS: (S)K43-LVMA)K-G-K23-STT-K23-(DMQ)K(LVMA)K-G-K23-(LVMA)K48-1A(DNR)K  
NAME: Ribosomal protein L13 signature  
CONSENSUS: (LVMA)K23VY(GK)K-M-LVMA)K(PT)K23-N-3H(KSR)K(QE)K(A)K-K23(LVMA)K44VY-  
CONSENSUS: (LPT)K-(G)K  
NAME: Ribosomal protein L14 signature  
CONSENSUS: (SA)K(LVY)K-KY-KS(DNR)K-G-K23-(PT)K-K23-(VT)K-K23-Y-LVMA  
NAME: Ribosomal protein L15 signature  
CONSENSUS: K-LVMAK23-KA(L)K23-(ST)K(LVMA)K-K23-3A(LVMA)K-(LVMA)K23-K-  
CONSENSUS: (LVMA)K-STT-K-K23-(LVMA)K-K23-K  
NAME: Ribosomal protein L16 signature 1  
CONSENSUS: (K)K-R-K(R)K(A)K(QV)K(LVMA)K-W-(LVMA)K(R)K(LVMA)K(LPT)K4P  
NAME: Ribosomal protein L16 signature 2  
CONSENSUS: R-K-G-K-(R)K-K-G-K23-(PV)K(R)

1033

1034

NAME: Ribosomal protein S11 signature.  
CONSENSUS: (LIVMF)<sub>2</sub>+GDTACT(LIVMF)<sub>2</sub>+C14QETALL+H<sub>2</sub>(1)HCDH(LIVMF)<sub>2</sub>+GLVMD.  
CONSENSUS: H(C)<sub>2</sub>(DEH)<sub>2</sub>+T-F+QPA(QTCH)<sub>2</sub>(DH).

NAME: Ribosomal protein S12 signature.  
CONSENSUS: (EK)<sub>2</sub>+P-P-LAAS)+R.

NAME: Ribosomal protein S13 signature.  
CONSENSUS: (ELQDS)<sub>2</sub>+R-H- $\alpha$ (2)-KSDH)<sub>2</sub>+C2(LIVMC)<sub>2</sub>R-G-Q.

NAME: Ribosomal protein S14 signature.  
CONSENSUS: (RP)<sub>2</sub>+H<sub>2</sub>(1)-C- $\alpha$ (1,12)(LIVMF)<sub>2</sub>+GLVMA)<sub>2</sub>(EC)<sub>2</sub>(RG) $\alpha$ (2)(RP).

NAME: Ribosomal protein S15 signature.  
CONSENSUS: (LIVM) $\alpha$ (2)+H(LIVMF)<sub>2</sub>+ $\alpha$ (2)-D- $\alpha$ (2)(SLAQ) $\alpha$ (2)(L)<sub>2</sub>+P $\alpha$ (9)(LIVM) $\alpha$ (2)-  
(PT).

NAME: Ribosomal protein S16 signature.  
CONSENSUS: (LIVMT)<sub>2</sub>+GLVMD(KL)<sub>2</sub>L-ETASQ-R- $\alpha$ (2)(AX).

NAME: Ribosomal protein S17 signature.  
CONSENSUS: G-D+GLV)+GLV)+ $\alpha$ (2)EK)<sub>2</sub>+P(K)<sub>2</sub>P(LV)<sub>2</sub>R.

NAME: Ribosomal protein S18 signature.  
CONSENSUS: (PT) $\alpha$ (PT)<sub>2</sub>+ $\alpha$ (2)(LIVMT) $\alpha$ (2)(LIVM) $\alpha$ (2)(PTT)<sub>2</sub>(LIVM) $\alpha$ (PT)<sub>2</sub>(DEP)<sub>2</sub>+  
CONSENSUS: (QTY)<sub>2</sub>(LIVM) $\alpha$ (2)+D-R-LVMAAS).

NAME: Ribosomal protein S19 signature.  
CONSENSUS: (ETP)<sub>2</sub>-G-ELB(QH) $\alpha$ (2)(LIVMD) $\alpha$ (2)(LIVM) $\alpha$ (2)(SD) $\alpha$ (2)(L)<sub>2</sub>(P) $\alpha$ (AS)<sub>2</sub>(DH)<sub>2</sub>P-  
CONSENSUS:  $\alpha$ (2)(PT).

NAME: Ribosomal protein S21 signature.  
CONSENSUS: (DH)<sub>2</sub>+ $\alpha$ (L)<sub>2</sub>(K)<sub>2</sub>R-P-E-(K)<sub>2</sub> $\alpha$ (2)(KL).

NAME: Ribosomal protein S24 signature.  
CONSENSUS: (LPT)<sub>2</sub>+SDH)+DP)<sub>2</sub>R+ $\alpha$ (2)<sub>2</sub>L+D-L.

NAME: Ribosomal protein S26 signature.  
CONSENSUS: H+ $\alpha$ (2)(LIVM) $\alpha$ (SA)<sub>2</sub>+P- $\alpha$ (2)W- $\alpha$ (LIVM) $\alpha$ (K).

NAME: Ribosomal protein S28 signature.  
CONSENSUS: (LIVM) $\alpha$ (TTAA)<sub>2</sub>+G+D+D- $\alpha$ (2)-G+P-M.

NAME: Ribosomal protein S29 signature.  
CONSENSUS: (KL)<sub>2</sub>L+R-S-L-R-E-S-P(LA)<sub>2</sub>+P(K).

NAME: Ribosomal protein S29 signature.  
CONSENSUS: R- $\alpha$ (2)-T-G(DA) $\alpha$ (2)(P)<sub>2</sub>R(K)<sub>2</sub>+R+ $\alpha$ (2)(LA)<sub>2</sub>G.

NAME: Ribosomal protein S30 signature.  
CONSENSUS: A-L(LSDP)<sub>2</sub>+Y-L- $\alpha$ (2)(SA) $\alpha$ (2)(DH)<sub>2</sub>G-L.

NAME: Ribosomal protein S37 signature.  
CONSENSUS: A+ $\alpha$ (2)(T)<sub>2</sub>R+L-L-R-R-(KL)<sub>2</sub>+ $\alpha$ (2)(PT)<sub>2</sub>+T-H.

NAME: Ribosomal protein S39 signature.  
CONSENSUS: P- $\alpha$ (2)(SA)<sub>2</sub>+C2(LIVMA)<sub>2</sub>+R+ $\alpha$ (LIV)<sub>2</sub>(LV)<sub>2</sub>Q+L(D).

NAME: Ribosomal protein S39 signature.  
CONSENSUS: L-T-V-P-R-E-C-S(LA).

NAME: Ribosomal protein S39 signature.  
CONSENSUS: (FA) $\alpha$ (Q) $\alpha$ (2)(L)<sub>2</sub>(STAA)<sub>2</sub>+G(PY) $\alpha$ (LA)<sub>2</sub>+GLVMD-Y-DH)<sub>2</sub>SD).

NAME: Ribosomal protein S39 signature.  
CONSENSUS: (TH)<sub>2</sub>C-V-S-C-A-H.

NAME: Ribosomal protein S39 signature.  
CONSENSUS: KDC) $\alpha$ (2)-C- $\alpha$ (2)-P- $\alpha$ (2)+P(A) $\alpha$ (2)-C- $\alpha$ (2)-C- $\alpha$ (2)-L- $\alpha$ (2)-P+G.

NAME: Ribosomal protein S39 signature.  
CONSENSUS: R-ET)<sub>2</sub>R-E-R+R+L.

NAME: DNA associated repeat protein mod. / mod. / PHE1 signature.

WO 81/12659

PCT/IB80/01196

CONSENSUS: Q P S G S A L.

NAME: DNA mismatch repair protein and family signature.  
CONSENSUS: (STY)G(LYM) + (LYM) + G S (LYM)Y(GC)(S R N)G(GT) + G G.

NAME: nucleic acid repair.  
CONSENSUS: G N T S E + (H) (TAC) (LYM)A C Y + S E (LYMPT) + S E.

NAME: DNA protein signature.  
CONSENSUS: (GCA)G(LYV)G (GSH)G + (H) (H R E) + H (STP) (STY) (LYM)G D + (S A) + (H) (H R E) (LYM).

NAME: Small, acid-soluble spore protein, alpha form type, signature 1.  
CONSENSUS: E + S (LYV) A + (DE) (LYMPT) G (LYMPT).

NAME: Small, acid-soluble spore protein, alpha form type, signature 2.  
CONSENSUS: (S L) (SAC) + G + Y G G + G (LYM) + (H R E) (LYM) G D.

NAME: Zinc-containing alcohol dehydrogenase signature.  
CONSENSUS: G N E G D G + (H) (G A) + (H) (Y S A C).

NAME: Oxidative acidification / iron-sulfur signature.  
CONSENSUS: (G D S) (G D G) + (H) A + (H) (S A) T G G + G + G + (H) G D (H R).

NAME: Iron-containing alcohol dehydrogenase signature 1.  
CONSENSUS: (T A L Y V) (LYV) + (DE) + (H) T P + (H) A (L Y V) + (G T) + (H) D (T A L Y M).  
CONSENSUS: (LYMPT) + (H) S.

NAME: Iron-containing alcohol dehydrogenase signature 2.  
CONSENSUS: (G T P) + (LYMPT) A C G (G R) + (H) (G A D) (G H Y G) + (LYMPT) (G A T T) (G A S) + (H) + (LYMPT) + (H) R G (G A) + (T T A C).

NAME: Short-chain dehydrogenase/reductase family signature.  
CONSENSUS: (LYMPT) A C G + (H) D Y (T T A C) C Y (T T A G H Q C Y V M) (T A G C) S. (H C) (S A G F R).  
CONSENSUS: (LYMPT) A C G + (H) (LYMPT) + (H) (LYMPT) Y M G A P T T Q G S A C Q E H M.

NAME: Aldo/keto reductase family signature 1.  
CONSENSUS: G (T P) + (H) G A L G (LYMPT) G (T A C) (A S) + (H) S + (H) D (LYM) G.

NAME: Aldo/keto reductase family signature 2.  
CONSENSUS: (LYMPT) + (H) (H R E Q) + (LYM) G (LYM) (G C) H (P V).

NAME: Aldo/keto reductase family signature without SH signature.  
CONSENSUS: (LYMPT) (P A V) (S A) (T T) + (H) S + (H) (G T A B Q E) (H L) + (H) (LYMPT) A L.

NAME: Haemolysin dehydrogenase signature.  
CONSENSUS: A + (H) G (LYMPT) (T A C) + (H) (H R E) P + (H) D (LYM) + G + G + (H) S.

NAME: NAD-dependent glyoxyl-3-phosphat dehydrogenase signature.  
CONSENSUS: G (A T) (LYM) S (D P) (LYMPT) A + (G A) + G (LYMPT) + (DE) G (LYM) + (LYMPT) G + S.

NAME: FAD-dependent glyoxyl-3-phosphat dehydrogenase signature 1.  
CONSENSUS: (P V) G G G G D G (T A C Y) G + A + G + (H) S G.

NAME: FAD-dependent glyoxyl-3-phosphat dehydrogenase signature 2.  
CONSENSUS: G C E G D (S T T E) + S + (H) A.

NAME: Monomer dehydrogenase signature.  
CONSENSUS: (LYMPT) + (P V) G D (T A C Y) + Y D S (P V) + (P S).

NAME: Flavinol dehydrogenase signature.  
CONSENSUS: S D + (H) A G P (T T) S (LYMPT) (LYM) A Y (A C) + (H) A + (H) (LYM) (A V).  
CONSENSUS: (S A C L) (DE) (LYMPT) (LYM) (S A) + (H) S M.

NAME: L-lactate dehydrogenase active site.  
CONSENSUS: (LYM) A + G (S Q) S G (D P) (S T).

NAME: Dimeric monomer 1 dehydrogenase NAD-binding signature.  
CONSENSUS: (LYM) A L (H) (P V) (LYMPT) + (H) + G + (H) (R E Q S A C) (LYM) G + (H) (S L A).  
CONSENSUS: (LYMPT) + (H) (P V) (C T R) (D N T E).

NAME: Dimeric monomer 2 dehydrogenase signature.  
CONSENSUS: (LYMPT) A L (LYMPT) + (H) (S A C) (H R E Q S A C) (LYM) A (LYMPT) + (LYMPT) (R O Q) +

PCT/D08/01496

1037

PCT/1230/01426

1038



PCT/TB00/01496

1039

1040

NAME: Manganese and iron superoxide dismutase signature.  
CONSENSUS: D-x-W-E-R-E-T-A-L-S-P-T-T-T.

NAME: B-Branched-chain reduction large subunit signature.  
CONSENSUS: W-x-D-L-F-P-x-A-7-y-G-L-V-M-Q-E-P-T-A-I-Q-N-Q-x-D-E-T-A-Q-L-V-M-Q-A-S-K-Q-x-D-  
CONSENSUS: [A].

NAME: B-Branched-chain reduction small subunit signature.  
CONSENSUS: [F-V-M-S-E-Q]-E-x-L-D-L-V-T-T-A-[D-V-T]-E-G-I-A-x-[E-T-A-V-M]-Y-x-D-L-V-M-Q-x-D-  
CONSENSUS: [L-P-T]-E-V-F-T-C-A-I.

NAME: Nitrogenase component 1 alpha and beta subunits signature 1.  
CONSENSUS: [L-V-M-F-T-T-Q]-[L-V-M-F-T-T]-H-A-Q-[A-G-S-P]-[L-V-M-M-Q-A]-[A-Q]-C.

NAME: Nitrogenase component 1 alpha and beta subunits signature 2.  
CONSENSUS: [E-T-A-N-Q]-[E-T-T]-C-x-D-G-D-D-[D-N-F]-[L-V-M-T]-E-[E-T-A-C-A]-[L-V-M-F-T-T].

NAME: NADH/ruC family signature 1.  
CONSENSUS: E-x-G-C-F-x-D-D-S-A-I-x-G-C-[A-Q]-G.

NAME: NADH/ruC family signature 2.  
CONSENSUS: D-x-L-L-G-D-V-V-C-G-G-P-[A-Q]-x-P.

NAME: Nickel-dependent hydrogenase large subunit signature 1.  
CONSENSUS: E-x-G-L-V-M-F-T-E-x-L-D-E-E-S-M-A-I-x-C-G-L-V-M-I-C.

NAME: Nickel-dependent hydrogenase large subunit signature 2.  
CONSENSUS: [F-T]-D-P-C-E-B-I-[A-S-E]-C-x-D-3-y-N.

NAME: Chaperonin-60 kDa subunit signature.  
CONSENSUS: H-L-V-Y-M-x-D-L-P-M-Q-[E-T-A-C]-[L-V-M-I]-[D-E-Q]-S-L-P-M-A-I-[L-V-M-Q-T]-[E-P]-E-  
CONSENSUS: -x-[Q-N]-[V-T]-[L-T]-[E-T-A-C]-Q-L-V-M-I-E-R-I.

NAME: Bacterial-type phytylase diacylglycerol signature.  
CONSENSUS: [H-Q]-x-[F-T-W-T]-[L-V-M-F]-x-G-[A-Q-C]-G-E-I-T-A-I-M-Q-T-P-G-[E-T-A-Y]-G-L-V-M-Q-  
CONSENSUS: x-D-E-S-L.

NAME: Glycine radical signature.  
CONSENSUS: [E-T-V]-x-E-[V-T-H-C-A]-G-T-x-[S-A-C-V].

NAME: Erythropoietin EPO/ERL24 family signature 1.  
CONSENSUS: D-x-D-L-P-N-Q-T-Q-x-[E-T]-x-G-x-D-A-I-P-P-E-L.

NAME: Erythropoietin EPO/ERL24 family signature 2.  
CONSENSUS: [L-V-M-Q]-D-H-E-x-D-R-D-x-D-C-x-D-E-T-O.

NAME: HMMT/PMPT/STMT family of aminotransferase signature.  
CONSENSUS: E-A-D-A-G-S-D-P-T-D-T-[V]-Q-A-L-L-S-A-I-C.

NAME: RNA methyltransferase track family signature 1.  
CONSENSUS: [D-N]-P-[P-A]-E-x-G-x-L-I-I-H-[L-V-M-E]-Y-x-E-C-H-x-D-T.

NAME: RNA methyltransferase track family signature 2.  
CONSENSUS: [L-V-M-Q]-D-x-P-P-[Q-T]-[E-T]-x-H-[L-V-M-F-T]-E.

NAME: Thymidine synthase active site.  
CONSENSUS: E-x-D-L-V-M-Q-x-D-[F-W]-[Q-N]-x-E-P-[L-V]-x-F-C-[L-V-M-Q]-x-D-[Q-M-T]-[F-T-W]-  
CONSENSUS: x-E-L-V.

NAME: Ribosomal RNA subunit diacylglycerol signature.  
CONSENSUS: [L-V-M-Q]-[L-V-M-F]-[E-T]-x-G-[A-Q]-x-[L-V-M-F]-[E-T]-x-D-L-V-M-Q-  
CONSENSUS: x-H-[L-V-M-T]-x-[E-T-A-G-T]-[L-V-M-F-T-M-C]-E-x-D.

NAME: Methylated DNA -guanine-cytosine methyltransferase active site.  
CONSENSUS: [L-V-M-F]-C-H-E-L-V-M-F-E-D.

NAME: H-4 Adenosine-specific DNA methylase signature.  
CONSENSUS: [L-V-M-A-C]-[L-V-F-T-W-A]-x-[D-N]-P-P-[F-T-W].

NAME: H-4 cytosine-specific DNA methylase signature.  
CONSENSUS: [L-V-M-F]-S-P-P-[F-T].

NAME: C-5 cytosine-specific DNA methylase active site.  
CONSENSUS: [D-E-N-E-S]-x-[L-V-L-V]-x-D-[E-T-C]-x-P-C-x-D-[F-V-L-E-Q]-I.

NAME: Cyt cytosine-specific DNA methylase C-terminal signature.  
CONSENSUS: [E]QGTTP-<D>-G-R-[ST]AQ-[LVNMF]-K-[L]VNM-<D>-[LVN]-K-[L]VNM.  
NAME: Protein-L-isoaspartyl methylase C-methyltransferase signature.  
CONSENSUS: [C]A-[D]-G-<D>-G-[PT]YV-<D>-[A]S-[P]-[PT]-(DP)->4.  
NAME: Unmethylated C-methyltransferase signature 1.  
CONSENSUS: [LVN]G[G]G[TA]L[G-P-G]-K-[L]VMP-[LVN]T-[LVN]M-[E]KMQG-[A]G.  
NAME: Unmethylated C-methyltransferase signature 2.  
CONSENSUS: V-<D>H-L-[D]-G-<D>-[P]YV-[G]S-<D>-[LVN]-K-[A]-[LVN]MPTWYAC-  
CONSENSUS: <G>LVNTP->4-G.  
NAME: siteCCQ2 methyltransferase family signature 1.  
CONSENSUS: Y-D->H-R-<D>-[LVN]-S-<D>-H-<D>-W.  
NAME: siteCCQ2 methyltransferase family signature 2.  
CONSENSUS: R-Y-[LVN]S-[PT]V-G-G-<[LVN]P-><D>-[LVN]-S->4.  
NAME: Serine hydroxymethyltransferase pyridoxal phosphate attachment site.  
CONSENSUS: [G]D-[LVN]MPT->[ET]MTP-[G]STT-[ET]E-[H]-R-[ET]L-[P]->G-[PAC]-[G]Q-  
CONSENSUS: [G]A-[G]A.  
NAME: Phosphoribosyltransferase transferase active site.  
CONSENSUS: G->[ET]M-[P]VT->[P]YVYQ[N]MA-[>-(DE)N]->[LVN]M]-D->G-<D>-[LVN]-  
CONSENSUS: <H>-[LVN].  
NAME: Arginine and creatine phosphotransferase signature.  
CONSENSUS: P->[E]S->S-[G]T]-S-7.  
NAME: Transketolase signature 1.  
CONSENSUS: R-<D>-[LVN]MPT->[E]MKG[ET]MELP-<L>-A-[E]N]-G-H-[LVN]MPT-[G]TA-><D>-  
CONSENSUS: [L]M->[G]S.  
NAME: Transketolase signature 2.  
CONSENSUS: G-[ET]QCS-[D]M-G-[P]A[EQ]-[ET]-[P]KQ->[P]ADM-[LVN]MTC-[DE]PTW-<D>-  
CONSENSUS: [ET]MPT-<D>-[D]GAL.  
NAME: Transketolase signature 3.  
CONSENSUS: [D]CNVMA-T-[ET]-M-P-[ET]A-[LVN]MPT.  
NAME: Transketolase active site.  
CONSENSUS: [LVN]M-<[LVN]G-E-[LVN]M-[P]AD->[ET]T->[DE]K[P]AQ-G-[LVN]->[A]DVS->  
CONSENSUS: [Q]E[ET]T->[LVN].  
NAME: Aspartate/lysine/tyrosine COT / CPT family signature 1.  
CONSENSUS: [L]E-P->[LVN]-P-[Q]VTA)-P->[LVN]->[D]KQ[A]S-[ET]([LVN]-<D>-[L]T).  
NAME: Aspartate/lysine/tyrosine COT / CPT family signature 2.  
CONSENSUS: S-[PT]W->[D]A-[E]A]-<E>-[LVN]MPT->[LVN]MPT->K-[D]HQ-[E]A]-<E>-  
CONSENSUS: [D]E-[P]E->[D]E-[G]A.  
NAME: Threonine xyl-transferase signature.  
CONSENSUS: [LVN]M-[N]T-><D>-C-[G]AGL-[ET]-[A]G-[LVN]MPTVH->[ET]AG-[LVN]-<E>-  
CONSENSUS: [LVN].  
NAME: Threonine signature 1.  
CONSENSUS: R-<D>-G-G->[LVN]-[A]->G-H-P->G->[ET]T-G.  
NAME: Threonine active site.  
CONSENSUS: [A]G-[LVN]M-[ET]AGLVN-[ET]AG-[LVN]M]-C->[A]G->[A]G->[A]G->[A]G->[A]G.  
NAME: Chalcone/flavone synthetase active site.  
CONSENSUS: Q-[LVN]M-R-[E]A]-<D>-D-G-[PT]E-H.  
NAME: Hexosephosphate-uracil succinyl transferase signature.  
CONSENSUS: [LVN]G[A]E[N]-K-[ET]AVI-[LVN]-K-[LVN]MPT-[LVN]G[A]E[N]-<D>-  
CONSENSUS: [ET]A-[V]S-[LVN]M-[E]A]-<D>-[ET]A-[V]S-[LVN]M-[LVN]MPT.  
NAME: Beta-1,4-galactosyl transferase active site.  
CONSENSUS: G-H-[LVN]MPTA]-<D>-[A]GQ-C-[ET]A-[ET]AG-[LVN]MPT.  
NAME: Chalcone and stilbene synthase active site.

WO 01/12659 PCT/IB00/01496

CONSENSUS: R-{LVMFYE}-x-{LVM}-x-{QHQ}-x-G-C-{FYNA}-[GA]-G-{GA}-[STAV]-x-{LVMFY}-  
 CONSENSUS: [RA].

NAME: Myristoyl-CoA protein N-myristoyltransferase signature 1.  
 CONSENSUS: E-I-N-F-L-C-A-H-E.

NAME: Myosin-CaA protein N-myristoyltransferase isoform 2.  
 CONSENSUS: K-F-G-L-G-D-G.

NAME: *Gemona glyoxylyltransferase* *gemona* *gemona*  
 CONSENSUS: T-(STA)-H-a-(ST)-(LIVMA)-a(a)-G-(SP)-a-V-(TA)-a-T-a-T-(LIVM)-(NE)-  
 CONSENSUS: a(1,2)-(FY)-G.

NAME: Transglutaminase active site.  
 CONSENSUS: 10T-Q-ICAF-W-Y-z-(3A)HGA-(TVT)-z(2)-T-z-(LMSC)-R-y(3A)H-LV-G

NAME: Phosphorylase pyridoxal-phosphate mechanism and  
 CONDENSED: E-A(3C)-G-z-(GS)-z-M-E-z(7)-(LM)-N.

NAME: UDP-glucosyltransferase signature  
 CONSENSUS: [FW]-x(2)-Q-x(2)-[LVMTA]-[LDMV]-x(4-6)-[LVGAC]-[LVPPA]-[LVMP]-[STAGC]-  
 CONSENSUS: [HMQ]-[STAGC]-G-x(2)-[STAG]-x(3)-[STAGL]-[LVMPA]-x(4)-[PQR]-[LVMT]-  
 CONSENSUS: x(3)-[PA]-x(3)-[DEG]-[QRMV]

NAME: *Paratyphimolus phagocytozoyl transferrin receptor*.  
 CONSENSUS: (L)YMFYVCTA(L)LYMG(L)YVMA(L)YVMPG(D)D(L)YVMS(L)YVM(L)STAYD(L)  
 CONSENSUS: (STAB)GAC(L)STAB(L)

NAME: *Chlamydomonas reinhardtii* step-1 gene  
 CONSENSUS: (FAS)-(LYMPTY)-(LYMPTY)-G-(LYMPTY)-C-(LYMPTYN)-G-z-(QDS)-z-(LYMFA)

NAME: Glycine synthetase class-II active site.  
 CONSENSUS: <sp(11)-C-IGS-IV-ILVMPYW-AG>

NAME: Perin and other phosphorylase family 1 signature.  
CONTENTS: KSTT-1-G-ILVMD-G-1(PA)-3-1-KSTAI-4-1-D-8-L

**NAME:** Purines and other phosphorylases family 1 signature.

NAME: Thyridine and pyrimidine nucleoside phosphorylase signature.  
CONSENSUS: S-PGD-R-CAI-E-LVY-GD-(TAI)GAI-G-T-A-D-A-[LVY]-E

NAME: ATP phosphoribosyltransferase regulators.  
 CONSENSUS: E-K-S-G-L-GLAG-L-T-V-L-D-LIV-L-T-ST-Q-L-T-L-M

NAME: NAD-arginine ADP-ribosyltransferase synthase.  
 CONDENSED: (FY)-s-(FY)-E-s(G)-H-(FY)-s-L-(ST)-s-A

NAME: Protoporphomycin dimethylglycyl transferase signature.  
 CONSENSUS: G-R-2 (GA)-N-P-FLYMGF-N-3-E-Q)-G

NAME: 8-adenosylmethionine synthetase signature 1.  
CONSENSUS: G-A-G-D-Q-G-N-G-Y.

NAME: S-adenosylmethionine synthetase signature 2.  
CONSENSUS: G-(GA)-G-(AC)-P-S-L-K-(DE).

NAME: Polypropyl syndiotactic signatus 1.  
 CONSENSUS: [L]V[Q]Z-1-D-D-2(2,4)-D-2(4)-R-R-[OH]

NAME: Polypyrrolyl synthetase signature 2.  
 CONSENSUS: [LVNMFY]-G-[L]-[FTL]-Q-[LVN]-I-Q-D-[LVNMFY]-I-[DNG]

NAME: Squirrels and phylogenetic synthesis signature 1.  
 CONSENSUS: Y4(GSAM)-x(2)-(V3Q)-A-(GSA)-(LIVAT)-(TV)-G-x(2)-(LMSC)-x(2)-(LIV)

NAME: Squares and polygons synthesis signature 2.  
CONSENSUS: [LVM]-Q-w(3)-Q-w(2,3)-M-[IF]-s-R-D-(LVMFY)-w(2)-(DE)-w(4,7)-R-s-(FY)

NAME: Protein preylransferases alpha subunit repeat signature.

**C. Optimization:**

NAME: *Radical synthesis alpha class family signature*  
CONSENSUS: [LVMF]-c7h-g(tstamq)-[kssq]w; y h [LVM]-s  
NAME: *Dihydroxyphenyl signature 1*  
CONSENSUS: [LVM]-s-[a]g-[LVMF]c7h-h-s-t-o-s-p-s-d-s-[g]  
NAME: *Dihydroxyphenyl signature 2*  
CONSENSUS: [c7h]s-a-[LVMKE]-d-[LVM]-g-[c7h]-w-[t]t-a-s-p  
NAME: *EPF signature 1*  
CONSENSUS: [LVM]-c7h-[c]h-s-[a]-g-t-[t]t-a-s-s-[LVMF]-s-[k]t-a  
NAME: *EPF signature 2*  
CONSENSUS: [t]s-s-s-[c]h-s-[t]t-[c]k]-s-[LVM]-[t]t-a-[LVM]-c7h-[c]h-[LVMF]-s  
CONSENSUS: [k]r-a-[LVMF]-g  
NAME: *PLAPKSTSLTCL family signature*  
CONSENSUS: Q-c7h-f-d-s-y-[t]t-s-s-s-c  
NAME: *Aminotransferase class-I pyridoxal-phosphate mechanism site*  
CONSENSUS: [c]h-[LVMF]t-a-[c]t-a-s-e-w-h-[s]a-l-y-h-[LVMF]-s-[k]r-a-s-s-s-[LVM]-s  
CONSENSUS: [g]a  
NAME: *Aminotransferase class-II pyridoxal-phosphate mechanism site*  
CONSENSUS: T-[LVMF]w-s-[t]a-g-s-[a]g-[LVMF]w-s-[a]g-c7h-[s]a-g  
NAME: *Aminotransferase class-III pyridoxal-phosphate mechanism site*  
CONSENSUS: [LVMF]w-c7h-s-d-s-[LVM]-c7h-[c]p-s-h-[LVMF]w-a-g-s-h-[s]a-c7h-s  
CONSENSUS: [s]a-s-h-c7h-s-h-s-[LVMF]w-c7h-s-[c]s-a-s-e-c7h-[t]t-a-w-[s]a  
NAME: *Aminotransferase class-IV signature*  
CONSENSUS: E-s-[t]a-g-c7h-s-[LVMF]c7h-[t]t-s-h-[LVMF]-s-t-h-s-[LVM]-s  
CONSENSUS: [c]h-[LVM]-s-[s]a-t-s-s-[LVMF]a-c  
NAME: *Aminotransferase class-V pyridoxal-phosphate mechanism site*  
CONSENSUS: [LVMF]c7h-[c]h-[LVMF]a-c-[LVMF]t-a-w-h-c7h-a-[c]t-a-[c]h-s  
CONSENSUS: -h-h-s-s-[s]a-t-s-s-[LVMF]a-c  
NAME: *Phosphorylation signature*  
CONSENSUS: [LVM]-g-p-[t]h-s-s-[t]p-f-c7h-[LVM]-[c]h-t-t-c7h-[LVM]-c7h-w-t-e-s  
CONSENSUS: [L]p  
NAME: *Catalytical signature*  
CONSENSUS: G-s-s-p-[LVM]-s-g-s-h-s-d-y  
NAME: *GTPase domain ATP binding domain*  
CONSENSUS: [LVM]-[p]h-s-[c]t-a-f-h-s-[g]-[c]h-s-s-[s]a-[c]t-a-c  
NAME: *Phosphotransferase signature*  
CONSENSUS: [s]a-s-h-c7h-s-s-[c]h-g-c7h-d-s  
NAME: *phb family of carboxylate kinase signature 1*  
CONSENSUS: [a]g-g-h-s-[s]a-f-s-h-s-[t]t-a-w-h-[c]h-h-y-g  
NAME: *phb family of carboxylate kinase signature 2*  
CONSENSUS: [p]h-s-[t]t-t-y-s-[a]g-s-[c]h-s-[a]g-y-[a]g-[LVMF]-[LVM]t-a-f  
NAME: *RCR family signature*  
CONSENSUS: [LVM]-c7h-g-[LVMF]c7h-g-s-[s]a-[LVMF]-h-s-g-c7h-s-[s]a-t-t-c7h  
CONSENSUS: g-[s]e  
NAME: *Phosphotransferase signature*  
CONSENSUS: E-[LVM]-s-s-d-c7h-g-s-[t]t-s-d  
NAME: *Threonine kinase cellular type signature*  
CONSENSUS: [k]a-c7h-[c]h-s-t-s-p-s-c-[h]e-s-s-c7h-[LVMF]w  
NAME: *POOT family of carboxylate kinase signature 1*  
CONSENSUS: [s]p-f-c7h-s-[t]t-c7h-s-[LVMF]w-s-w-[LVMF]-s-[c]h-t-t-s-[c]h-s  
NAME: *POOT family of carboxylate kinase signature 2*  
CONSENSUS: [c]s-a-s-[LVMF]w-s-g-[LVM]-c7h-[c]h-[c]h-s-[LVMF]-c7h-s-[t]t-a-s  
CONSENSUS: s-[LVMF]c7h-g

NAME: Protein kinase ATF-binding region signature.  
CONSENSUS: [LPT]G-[PT]G-[PT]PTFMGQTNQ[RGAA]P[W]L[VCAT]Y[PD]HGGTACLVNMFY.  
CONSENSUS: K21H-LVYMFYCTCTA[3]A[3]A[3]Y[3]L[3]V[3]M[3]A[3]C[3]E[3]A[3]E.  
NAME: Serine/threonine protein kinase active site signature.  
CONSENSUS: [LVNMFPTC]+[KT]+D-[LVNMFY]+E-W[H]-N-[LVNMFYCTT3].  
NAME: Tyrosine protein kinase specific active site signature.  
CONSENSUS: [LVNMFPTC]+[KT]+D-[LVNMFY]+[KSTAC]-K21H-[LVNMFYCTE].  
NAME: Protein kinase domain profile.  
NAME: Casein kinase II regulatory subunit signature.  
CONSENSUS: C-P-K-[LVNMFY]+C-K21H-P-[LVNMFAC]G-K21H-V[KKR]-K21H-C-P-K-C.  
NAME: Pyruvate kinase active site signature.  
CONSENSUS: [LVYACQ]+[LVNMFED]S[AFCV]E-[LVY]E-[HKEST]+[DEQRH]GST[A]L[LVNMF].  
NAME: Shikimate kinase signature.  
CONSENSUS: [KRS]-K21H-E-[LVNMF]-K21H-[LVNMFED]S[A]-G21H-[LVNMF].  
NAME: Prokaryotic deoxyribonucleic kinase signature.  
CONSENSUS: S-L[LVNMF]-N-[STT]S[A]-L[LVY]-S-K21H-V-D.  
NAME: Phosphatidylinositol 3- and 4-kinase signature 1.  
CONSENSUS: [LVNMFAC]E-K21H-[DEA]GSD-[LVNMFAC]E-Q[DE]-K21H-Q.  
NAME: Phosphatidylinositol 3- and 4-kinase signature 2.  
CONSENSUS: [KSD]+[A-Y]-K21H-[LVNMF]-K21H-[FTT]L[LVNMFED]S[A]-[LVNMF]+D-E-H-K21H.  
NAME: Arsenate and butyrate kinase family signature 1.  
CONSENSUS: [LVNMFED]+[LVNMF]-N-K21H-[STT]S-[K21H].  
NAME: Arsenate and butyrate kinase family signature 2.  
CONSENSUS: [LVNMFED]-K21H-K21H-G-[STT]L[LVNMF]+[A-Y]-K21H-G.  
NAME: Phosphoglycerate kinase signature.  
CONSENSUS: [EAKGTCVYPT]L[LVNMF]L[LVNMFAC]S-E-Q-K21H-[SACVY]-P.  
NAME: Aspartokinase signature.  
CONSENSUS: [LVNMF]+E-[PTT]G-D-[STT]K21H-[LVNMF].  
NAME: Chemoattractant 3-kinase signature.  
CONSENSUS: [GTTN]-K21H-G-G-[KCK]S[M]+[STTA]E-[LVNMF]+[SAH]TCAH-K21H[SALV].  
CONSENSUS: K21H-G.  
NAME: ATP gamma phosphate/adenosine active site.  
CONSENSUS: C-P-K21H-[STT]H-[K21H]-G-T.  
NAME: PTD HPS aspartate kinase phosphorylation site signature.  
CONSENSUS: G-[LVNMF]-H-[STTA]E-[PA]G21H-[STTA].  
NAME: PTD HPS aspartate kinase phosphorylation site signature.  
CONSENSUS: [K21H]E[DEKCTVY]-K21H-[LVNMFED]S[A]-[LVNMF]+[LVNMF]-P-N-[KAC].  
NAME: PTD EEA domain phosphorylation site signature 1.  
CONSENSUS: G-K21H-[LVNMFED]H-[LVNMF]-G-[LVNMF]+T-[A-LV].  
NAME: PTD EEA domain phosphorylation site signature 2.  
CONSENSUS: [EPTQD]-K21H-[LVNMF]S[A]-K21H-[LVNMF]-A-[LVNMF]-P-N-[KAC].  
NAME: PTD EER domain tyrosine phosphorylation site signature.  
CONSENSUS: H-[LVNMFPT]-K21H-C-K21H-[LVNMF]-L-[LVNMF]+[LVNMF]+[EDQ].  
NAME: Adenylate kinase signature.  
CONSENSUS: [LVNMFPT]W21H-D-G-[PTG]-P-K-K21H-[HQS].  
NAME: Nucleotide diphosphate kinase active site.  
CONSENSUS: H-K21H-[SA]-S-D-[SA]L[LVNMFCKE].  
NAME: Chomplex kinase signature.  
CONSENSUS: T-[PTT]-E-K21H-[K21H]-K21H-[DE]-K21H-G-K21H-Y-K21H-[PTT]-L[LVNMF].

NAME: *Quercus laevis* domain profile.  
NAME: Phosphatase/ pyrophosphatase signature.  
CONSENSUS: D-L-S-R-(S-A)-Q-(H-D-T)K-D-Q-(P-T)P-K-D-P-(L-V-M-P)D.  
NAME: T 8-ribonucleic hydrazinylphosphatase/pyrophosphatase signature.  
CONSENSUS: G-L-P-E-S-K-D-Q-L-D-L-L-V-M-E-D.  
NAME: Bacteriophage spx RNA polymerase family active site signature 1.  
CONSENSUS: P-(L-V-M)-K-D-D-(S-A)-(P-T) (A-C)-(D-H)-(S-A)-(L-V-M-P-T)-Q.  
NAME: Bacteriophage spx RNA polymerase family active site signature 2.  
CONSENSUS: (L-V-M-P-T)-R-K-K-D-S-K-K-D-L-V-M-M-M-(P-T)-K-D-T.  
NAME: Eukaryotic RNA polymerase II hypophosphate region.  
CONSENSUS: Y-(P-T)-P-(P-T)-S-P-(S-T-A-M-E).  
NAME: RNA polymerase beta chain signature.  
CONSENSUS: Q-K-E-(L-V-M-P-A)-(P-T-A-C)-(S-T-N)-Q-(S-T-A)-K-D-Q-Q-M-H-K-Q-(P-V-T).  
NAME: RNA polymerase M / 13 kD subunit signature.  
CONSENSUS: P-C-S-(P-E-K-S-T-T)-C-(G-N-E)-(D-H-S-A)-(L-V-M-M-Q)-(L-V-M)-M-S-I-H-C-K-D-C.  
NAME: RNA polymerase D / 30 to 40 kD subunit signature.  
CONSENSUS: H-(S-A)-(L-V-M-P)-S-K-M-T-(S-A)-M-T-Y-M-G-N-K-(P-T-A)-K-D-(D-H)-S-K-(L-S).  
CONSENSUS: (Q-A)-S-K-S-(G-A)-(L-V-M-E)-P.  
NAME: RNA polymerase H / 33 kD subunit signature.  
CONSENSUS: H-(M-E)-(L-V-M-Q)-Y-P-K-N-K-D-(L-V-M)-K-D-(D-E).  
NAME: RNA polymerase E / 34 to 35 kD subunit signature.  
CONSENSUS: (P-T)-S-(P-T)-K-S-K-M-T-S-K-S-D-(P-M)-(S-A)-S-K-(S-A)-Q.  
NAME: RNA polymerase L / 33 to 35 kD subunit signature.  
CONSENSUS: (D-M-E)-H-(P-T)-(L-V-M)-Q-(A-P)-H-K-(I)-Y-S-(P-M)-M-D-Y-M-(S)-M-P.  
NAME: RNA polymerase H / 8 kD subunit signature.  
CONSENSUS: (L-V-M-M-E)-P-(L-V-M-Q)-K-C-P-(P-T)-C-Q.  
NAME: DNA polymerase family A signature.  
CONSENSUS: R-K-D-(G-S-A-Y)-S-K-(L-V-M-P-T)-(A-Q)-K-D-Y-K-D-(D-S)-K-D-(L-V-M-A-L).  
NAME: DNA polymerase family B signature.  
CONSENSUS: (T-A)-(G-L-V-M-E-T-A-C)-D-T-D-(S-D)-(L-V-M-P-T-C)-(L-V-M-T-A-C).  
NAME: DNA polymerase family B signature.  
CONSENSUS: G-(S-D)-(L-V-T)-S-K-(G-S-K)-K-(H)-(S-U-C-L)-D-(L-V-M)-D-(L-V-M-P-T)-D-K-D-(S-A-P).  
NAME: Quotase-1 (phosphate or/and translocase family 1) active site signature.  
CONSENSUS: P-E-N-(S-E)-G-K-D-G-M-T-H-P-H-K-Q.  
NAME: Quotase-1 (phosphate or/and translocase family 2) signature.  
CONSENSUS: D-L-P-A-Y-K-G-(P-T)-(L-V-M-E)-D-S-A-N-(D-E-M)-H-(P-T)-Q-G-G.  
NAME: ADP-glucose pyrophosphatase signature 1.  
CONSENSUS: (A-G)-G-G-S-G-(P-T-K)-L-K-D-L-(T-A)-K-D-A-S-P-A-(L-V).  
NAME: ADP-glucose pyrophosphatase signature 2.  
CONSENSUS: W-(P-T)-Q-(P-T)-A-(D-N-G-L)-A-(L-V-M-P-T-T-W).  
NAME: ADP-glucose pyrophosphatase signature 3.  
CONSENSUS: (A-P-T)-G-S-M-G-(L-V-M-M)-Y-(P-V-C)-(L-V-M-P-T)-K-D-(D-E-N-P-M-E).  
NAME: Phosphatase cytidyltransferase signature.  
CONSENSUS: S-(L-V-M-P)-S-K-M-T-K-D-S-(S-A)-K-D-(L-H)-(P-C)-M-G-G-(L-V-M)-D-D-S.  
CONSENSUS: (L-V-M-P-T)-D.  
NAME: Ribonuclease PH signature.  
CONSENSUS: C-(D-E)-(L-V-M-E)-Q-(S-T-A)-D-G-(S-K)-K-D-(T-A)-A.  
NAME: 3'-5'-oligoribonucleic apurinic signature 1.  
CONSENSUS: G-G-S-A-(S-A)-(S-A)-T-T-A-L-E-E-(S-T)-S-A-D-(A-G).  
NAME: 3'-5'-oligoribonucleic apurinic signature 2.



1047

PCT/IB02/01496

1048

PCT/IB00/01496

1049

WO 01/12659

PCT/IB00/01496

CONSENSUS: F-G-G-R-F-E-Y-a-W-D-Y.  
NAME: Tetrahelix segment 2.  
CONSENSUS: Q-W-D-a-P-a-G(A)-W-(P-A)-P.  
NAME: Alpha-L-farnesylthio penicillin active site.  
CONSENSUS: P-KZD-a-KZ-E-W-S-a-C.  
NAME: Glyoxyl hydrolase family 1 active site.  
CONSENSUS: (L-V-M-F-T-T-C-L-V-P-F-F-E-G-L-V-M-F-T) S-H-G-(L-V-M-F-A-R)-C(SA-G).  
NAME: Glyoxyl hydrolase family 18 internal segment.  
CONSENSUS: F-a-P-Y-W-M-Q-(G-T-A)-a-(G-T-A)-a-(G-T-A-E-D)-F-Y-M-H-(M-Q)-a-E-a-(G-T-A).  
NAME: Glyoxyl hydrolase family 2 segment 1.  
CONSENSUS: H-a-(L-V-M-F-T-T-Q-D)-S-(G-T-A-C-M-E)-Y-P-H-a-(L-V-M-F-T-T-Q-D)-a-(D-W)-a-D-  
CONSENSUS: D-L-V-M-F-T-T-Q-D.  
NAME: Glyoxyl hydrolase family 2 acylthioase catalytic.  
CONSENSUS: (D-E-M-Q-P)-L-K-V-W)-H-H-L-A-P-(S-A-C)-(L-V-M-F-S)-W-K-D-Q-a-L-L-T-H-H-S.  
NAME: Glyoxyl hydrolase family 3 active site.  
CONSENSUS: (L-P-M-K-D)-(E-S)-a-(D-Q)-a-L-a-D-L-V-M-F-T-(L-V-T)-(L-V-M-F)-D-T-T-D-a-D-  
CONSENSUS: (S-G-A-D-W).  
NAME: Glyoxyl hydrolase family 3 segment.  
CONSENSUS: (L-P)-L-V-M-F-T-W-S-A-D-D-(D-M-Q-D)-(L-V-M-Q-T)-H-S-Q-P-Y)-(E-M-H-T-L-V-F-T).  
NAME: Glyoxyl hydrolase family 4 segment 1.  
CONSENSUS: V-a-Y-a-D-P-a-S-D-C-(S-A-P)-a-D-(G-S-A-D)-a-D.  
NAME: Glyoxyl hydrolase family 4 segment 2.  
CONSENSUS: (L-V-M-T-A)-(L-V-T)-(L-V-T)-(L-V-T)-S-P-D-(S-A-L)-(L-H)-P(S-A-C).  
NAME: Glyoxyl hydrolase family 5 segment.  
CONSENSUS: A-(G-T)-D-(A-C)-D-a-D-(D-M)-H-a-(S-A)-(L-V-M)-(L-V-M-Q)-a-A-a-L-(P-W).  
NAME: Glyoxyl hydrolase family 9 active site segment 1.  
CONSENSUS: (T-T-Y)-a-(L-V-M-F-T)-(G-T-T)-a-D-a-(H-E)-a-D-(L-V-M)-H-a-S.  
NAME: Glyoxyl hydrolase family 9 active site segment 2.  
CONSENSUS: (P-T-W)-a-D-a-L-V-T-T-Q-D-a-D-a-(G-T-A)-a-D-H-(G-T-A).  
NAME: Glyoxyl hydrolase family 10 active site.  
CONSENSUS: (K-T-A)-a-D-(L-V-W)-a-(P-M-F)-(G-T)-S-(L-T)-(D-W)-(L-V-M-F).  
NAME: Glyoxyl hydrolase family 11 active site segment 1.  
CONSENSUS: (P-M-A)-(L-E)-a-D-Y-T-(L-V-M-Q-D)-(D-E)-a-P-Y-W-H.  
NAME: Glyoxyl hydrolase family 11 active site segment 2.  
CONSENSUS: (L-V-M-F)-a-D-E-(A-G)-(T-W-Q)-(D-P-Q)-(S-D)-(G-T-A)-H-Q-a-(S-A-P).  
NAME: Glyoxyl hydrolase family 14 active site.  
CONSENSUS: S-(L-P)-D-(L-V-T)-a-D-L-E-a-D-(D-Q)-(L-S-H-P)-a-P(T-A).  
NAME: Glyoxyl hydrolase family 17 segment.  
CONSENSUS: (L-V-M)-L-(L-V-M-F-T)-W-A-L-E-D-T-A-Q-S-(G-T-A)-G-W-P-(T-T)-a-(S-A-Q).  
NAME: Glyoxyl hydrolase family 21 active site segment.  
CONSENSUS: D-(L-V-M)-a-D-(M-Q)-(P-Q)-a-Y-H-Q-a-L-V-M-F-T-(T)-E-a-(G-T)-S-(D)-a-D-  
CONSENSUS: Y-a-D-W.  
NAME: Glyoxyl hydrolase family 21 active site.  
CONSENSUS: (K-P)-(L-V-M-F)-W-a-D-M-(S-A)-E.  
NAME: Glyoxyl hydrolase family 21 segment 2.  
CONSENSUS: G-(A-Y)-D-(L-V-M-T)-C-a-D-T-T-a-D-(G-T)-a-D-L-C-a-S-W-a-D-(L-V)-(D)-(S-A)-  
CONSENSUS: F-a-P-P-a-S-(D-W).  
NAME: Glyoxyl hydrolase family 22 active site.  
CONSENSUS: H-a-D-P-a-L-(L-V-M)-H-G-P-H-G.  
NAME: Glyoxyl hydrolase family 22 penicillin active site.  
CONSENSUS: G-G-P-(L-V-M)-D-a-D-Q-a-S-H-S-P-T).

PCT/IB06/01496

1051

WO 01/2659

PCT/00091/96

NAME: Serum protein, Y1 family, tandem repeat int.  
 CONSENSUS: [STT]G(LVNPFTY)3(GPT)KZT(LVNA)+T(KZ)H.  
 NAME: Serum protein, Y1 family, serum repeat int.  
 CONSENSUS: T(KZ)KCH(MQ)S-G-S+(LVN)(PT).  
 NAME: Serum protein, serum family signature 1.  
 CONSENSUS: W T D A S A P A T.  
 NAME: Serum protein, serum family signature 2.  
 CONSENSUS: A-G-Y-Q-S(GT)S-(PT)W-S-(PT)Y-(TH)A+-G-G(GT)Y.  
 NAME: Putative endoplasmic family serum repeat int.  
 CONSENSUS: D-KZ+-KZT(LVNPFTY)K(LG-Q-S-S-G(LVNPFTY)S).  
 NAME: Endoplasmic Ctp serum repeat int.  
 CONSENSUS: T(KZ)LVNPFTY-G-A-SAC(S)-P(SA)STAL.  
 NAME: Endoplasmic Ctp tandem repeat int.  
 CONSENSUS: S-KZ(LAP)KZ(LVNPFTY)-H(LVNA)H-Q-P.  
 NAME: ATP dependent serum protein, but family, serum repeat int.  
 CONSENSUS: D-G-PT(S)S-A-SGS(LVNA)A(TA)SLPTM.  
 NAME: Endoplasmic dual (cytosol) protein serum repeat int.  
 CONSENSUS: Q-KZ(KGE)+C-(TW)KZ(STADCK)STADCK.  
 NAME: Endoplasmic dual (cytosol) protein tandem repeat int.  
 CONSENSUS: (LVN)KST(A)S-H-(G)SAC(S)LVNA)+LVNMA(T)D+-(G)SAD(H).  
 NAME: Endoplasmic dual (cytosol) protein signature repeat int.  
 CONSENSUS: (PTCH)(W)(LVY)+ELRQAG(H)STT)W-KZ(LFY)G-KZ-G(LFYW)  
 CONSENSUS: (LVNPFTY)+(LVN)P.  
 NAME: Uniquitin carboxyl-terminal hydrolase family 1 cytosol repeat int.  
 CONSENSUS: Q-KZ)H(SA)C-Q-KZ(LVNA)S(H)SA(LVNA)SA.  
 NAME: Uniquitin carboxyl-terminal hydrolase family 2 signature 1.  
 CONSENSUS: G-LVNPFTY)K(L)S(GC)PLAS(H)+C-(PT)W(LVNA)K(PST)SAC(Y)+LVNMS.  
 CONSENSUS: Q.  
 NAME: Uniquitin carboxyl-terminal hydrolase family 2 signature 2.  
 CONSENSUS: Y-L-L-S(SAG)(LVNPFTY)KZ(H-S-G-H)S-G-H-Y.  
 NAME: Cytosol family tandem repeat int.  
 CONSENSUS: H-KZ(LVNA)KZ(LVNPFTY)2-STT)H-G.  
 NAME: Cytosol family cytosol repeat int.  
 CONSENSUS: E-P-S(LVNPFTY)Q-A-C(SQG)G.  
 NAME: Endoplasmic and vent repeat protein repeat int.  
 CONSENSUS: (LVN)MSA(LVNA)T(A)S(LVNA)D-STT)G-ST(A)V(STA)P(SH)S+(LVN)P(SH)S  
 CONSENSUS: +(LVN)P(SH)S.  
 NAME: Nuclear and mitochondrial, one binding repeat signature.  
 CONSENSUS: (STAL)LVNA)KZ)H-S(LVNPFTY) (DE)KST)H+-LVN)PFTY)G(SH).  
 NAME: Mitochondrial cytosol repeat.  
 CONSENSUS: P-S-C(SD)S-H(DS)ELVSA(PKQ).  
 NAME: Endoplasmic family, one binding repeat signature.  
 CONSENSUS: G-KZ)H-G-(STT)H-LVNPFTY(LVNA)KZ(DERN)DEK(L)SAG(AT)S-LVNPFTY)+  
 CONSENSUS: KST)H)KST).  
 #  
 AC: PRB14;  
 DE: Oxytocin family signature.  
 CONSENSUS: (LS)S(LA)T)KZ(LVNPFTY)2(DQ)K(L)+P+(LVNPFTY)KZ(H-KZ)SAG(H)  
 CONSENSUS: (LVN).  
 NAME: Proteasome A type subunit signature.  
 CONSENSUS: (PT)KZ(LVNPFTY)+(PT)W)S-P+G(KSL)KZ-Q(LVNA)DE(Y)S(LS)KZ)  
 CONSENSUS: (SAG).

1053

WD#U12A59

PCT/BD0091496

CONSENSUS: [SA]([LVN])(MGS)(TA)-D-D-F.  
NAME: Cytidine and deoxycytidine dinucleotide anti binding repeat sequence.  
CONSENSUS: [CDS]A(AD)YK-K(ND)(LVN)SA(TY)(LVN)-[T]LH-P-C-KD-E-C-K(T)(LVN).  
NAME: GTP synthetase domain 1.  
CONSENSUS: [DS](LVN)EY-K(D)(KRN)(DNI)(LVN)-K(T)ST-C-C-S-M-N.  
NAME: GTP synthetase domain 2.  
CONSENSUS: [SA]+(SA)+Q(LVN)Q-S(RN)(L)(TFR).  
NAME: Histidine / tyrosine hydrolase signature 1.  
CONSENSUS: [G-K](LVN)FFYD+(TF)+S-K(D)(LVN)-Q-T-P.  
NAME: Histidine / tyrosine hydrolase active site signature.  
CONSENSUS: Q-GAQD-K(D)-C(Y-A)-S-P(H)-K(D)(PST)(LVN)FYK+(K).  
NAME: Intergate pyrophosphatase signature.  
CONSENSUS: D-K(D)D-D-P(D)(LVN)F-D(LVN)GAC.  
NAME: Acetylcholinesterase signature 1.  
CONSENSUS: [LVN]-G-Y-Q-G-Y-G-TH-S.  
NAME: Acetylcholinesterase signature 2.  
CONSENSUS: G-P(YN)A(YC)(L)Q(L)M)N-K(D)-G-Y-K(D).  
NAME: ATP synthase alpha and beta subunits signature.  
CONSENSUS: F(SAP)(LVN)(DNR)-K(D)-S-S-S.  
NAME: ATP synthase gamma subunit signature.  
CONSENSUS: [VY]T-S-S-K(KD)(K(D)-G-A-(SAR).  
NAME: ATP synthase delta (DECF) subunit signature.  
CONSENSUS: [LVN]+(LVN)PTT-K(D)(LVN)(DNRK)-K(D)(LVN)+(SA)-Q(LVN)PTGA-  
+([LVN)(L)EKEKQ-S(KSE).  
NAME: ATP synthase epsilon subunit signature.  
CONSENSUS: [ST]MNR+(ST)G(LVN)F-S-L-(SACY)N(LVN)T.  
NAME: ATP synthase epsilon subunit signature.  
CONSENSUS: [G]ETAS-S-P(Q)-F-K(D)(LVN)FFYK(D)(LVN)FFYK+(DEL).  
NAME: S1-E2 ATPase phosphorylation site.  
CONSENSUS: D-N-T-G-L(SR-TH).  
NAME: Serine and proline ATPase beta subunit signature 1.  
CONSENSUS: [PTW]-K(D)(PTW)+(PTW)(DNI)N-S(LVN)-G-S-T-K(W).  
NAME: Serine and proline ATPase beta subunit signature 2.  
CONSENSUS: [SA]-K(D)-C(KS)W)-K(D)-L-K(D)-G(A)-G.  
NAME: GDA(KD) family of nucleotide phosphatase signature.  
CONSENSUS: [LVN)+G-K(S)-G-(PT)+(PTW)(LVN)(TAG)+N-(PT).  
NAME: Indolylamine decarboxylase active site.  
CONSENSUS: S-P-L-Y-S-P-G-S-E-A)-T-C-P-P.  
NAME: Cytosine, active site site.  
CONSENSUS: P-S(TA)-G(LV)(VY)+(DS)-G-Y-S(Q)A-G.  
NAME: Cytosine, sequence and histidine active site.  
CONSENSUS: C-K(D)-S-(PT)-C-G-KSTY-K(D)(LVN)-K(D)-H-M.  
NAME: DDC / GAD / IDC / TyDC pyridoxal-phosphate dependent site.  
CONSENSUS: S-K(LVN)Y-K(S)-S(LVN)FFYK(D)(LVN)FFYK+(K)-K(D)(LVN)FFYK.  
CONSENSUS: K(D)-K(S).  
NAME: Ornithine decarboxylase family 1 pyridoxal-P dependent site.  
CONSENSUS: [T]A(Y)-S-S-S-K(D)(ST)A(KD)-ST(A)-Q-ST(A)E.  
NAME: Ornithine decarboxylase family 2 pyridoxal-P dependent site.  
CONSENSUS: [PT](FA)-S-S(SAC)(D)CLPW)-K(D)(LVN)F(LVN)TA)-K(D)(LVN)A)-K(D)-  
K(D).



1055

1056

CONSENSUS: [C]A[3]L[L]V[M]G[L]V[M]P[F]Y[2]-G-[G]T[3]T[3]G-G-S-[G]A[3]P[F]Y-K[N]-E[3]Q.

NAME: Dihydrodipicolinate synthetase, segment 2.  
CONSENSUS: Y-C[N]K[L]L[V]M[F]P-Y-K[2]E[3]T[3]-K[3]-[L]V[M]G-K[1,1,1]-[L]V[M]G-Y-[G]A[3]-[L]V[M]P.  
CONSENSUS: E-[D]E[3]Q[F]Y-[G]T[3]A[3].

NAME: RNAi family of penicillinase synthase segment.  
CONSENSUS: G-S-L-D-K[2]E[3]T[3]-G-G-[L]V[M]P[F]Y-K[2]T[3]-[D]H[T].

NAME: Cytochrome synthetase/penicillinase-like cytochrome P-450 synthase segment one.  
CONSENSUS: E-K-E-K[3]P[3]A[3]-[G]T[3]A[3]G[3]-G-[L]V[M]P[F]Y-K-G-[G]A[3]-K-E-B-K-[G]T[3]A[3]K[2]-[L]V[M]G.

NAME: Penicillinase and tetracycline resistance synthase segment.  
CONSENSUS: G-[G]T[3]G[L]V[M]G-[G]T[3]G-[K]C[3]-S-G-[D]H[1]-L-K-P-L-[G]A[3]-K[2]-[H]A[3].

NAME: Penicillinase synthetase, surface binding site.  
CONSENSUS: E-S-K-[L]V[M]P[F]A[3]-K[2]L[L]V[M]P[F]Y-G-[G]A[3]-C-K-[P]V[T]P-[L]V[M]P[F]G[3]A[3].

NAME: Cytochrome oxidase, penicillinase synthetase segment one.  
CONSENSUS: [D]Q[L]V[M]P[F]Y-K[2]H-[G]T[3]A[3]G[3]-[G]T[3]A[3]G[3]-T-S-[P]V[M]G[L]V[M]P[F]Y-K-G-[K]Q[2]-[D]H[N].

NAME: Cytochrome c and c1 haem synthase segment 1.  
CONSENSUS: [H]Q[L]V[M]P[F]Y-L[L]V[M]P[F]Y-K[2]P[3]-[G]T[3]A[3]-K[2]P-[T]M-K[2]H-[L]A[3]P[3]-G-[L]A[3]P[3].

NAME: Cytochrome c segment 1.  
CONSENSUS: G-[P]T[3]G[2]-H[L]H-[G]A[3]-[L]V[M]P[F]Y-K[N]H-[H-[P]V[M]P[F]Y-K[2]A[3]-[G]T[3]A[3]K[2]-[D]H[N].

NAME: Cytochrome c and c1 haem synthase segment 1.  
CONSENSUS: H-H-K[2]P[3]-S-K[2]-W-[K]Q[2]E[3]-K[2]-W-E.

NAME: Cytochrome c and c1 haem synthase segment 2.  
CONSENSUS: P-S-G-S-[D]-W.

NAME: Adenylate cyclase, class I segment 1.  
CONSENSUS: S-T-P-G-[S]A[3]T[3]-L-W-K-L-T-E.

NAME: Adenylate cyclase, class I segment 2.  
CONSENSUS: T-E-P-K-W-[D]Q[2]-[L]V[M]G-K-T-L-E-P-K-G.

NAME: Cytochrome c synthase segment.  
CONSENSUS: D-V-[L]V[M]G-K[2]L-G-K[2]H-[P]T[3]-[L]V[M]G-[P]V[M]P[F]Y-K[2]G-[D]H[N]E[3]W[3]-[D]H[T]E[3]V[3].  
CONSENSUS: [D]H[T]A[3]-K[2]D[3]E[3].

NAME: Chemosensory segment 1.  
CONSENSUS: G-E-S-H-[P]C[3]-K[2]L[L]V[M]G-[P]T[3]-[L]V[M]G[2]-[D]E[3]-G-K-[P]V[3].

NAME: Chemosensory segment 2.  
CONSENSUS: [K]E[3]-E-[G]A[3]G[3]-[G]A[3]G[3]-[P]V[3]-[G]T[3]-K[2]-[S]H[3]-Y-K[2]G[3].

NAME: Chemosensory segment 3.  
CONSENSUS: S-[E]H[3]-G-[P]V[3]-[G]A[3]V[3]-K[2]H[S]A[3]-[P]V[3]P[F]Y-[L]V[M]G-Y-[G]T[3]A[3]H[3]-[L]V[M]G.

NAME: 6-pyruvoyl tetrahydrophosphate synthase segment 1.  
CONSENSUS: C-N-S-K[2]G[3]-G-N-G-N-N-Y.

NAME: 6-pyruvoyl tetrahydrophosphate synthase segment 2.  
CONSENSUS: D-N-S-K[2]-N-L-D-N-D.

NAME: Penicillinase segment.  
CONSENSUS: [L]V[M]P[F]Y-K[2]-S-K-[G]S-[L]V[M]G-P-K[2]H-[D]E[3]Q[2]E[3]-K-G-D-K-Y.

NAME: Abiose synthase, penicillinase synthetase site.  
CONSENSUS: Y-K-E-K-[D]H[T]G[3]A[3]-Y-G-K-G.

NAME: Arginine and glutamine transaminase segment 1.  
CONSENSUS: [P]V[3]-[L]V[M]G-Y-C-K[2]L[3]-S-[G]T[3]-[G]A[3]-[G]T[3]G-[L]V[M]P[F]Y-T-T-A[N]K[3].

NAME: Arginine and glutamine transaminase segment 2.  
CONSENSUS: [L]V[M]G[2]-Y-[G]A[3]G[3]-C-T-[G]E[3]G[3]-[L]V[M]P[F]Y-[P]V[M]G[3]-[L]V[M]G.

NAME: Mandelate racemase / mandelate transaminase synthase family segment 1.  
CONSENSUS: A-K-[S]A[3]G[3]-[L]V[M]G-[D]E[3]-Y-A-K[2]-D-K[2]-[G]A[3]-[H]A[3].

NAME: Mandelate racemase / mandelate transaminase synthase family segment 2.

CONSENSUS: G-KT-D-WP-A-K(14)LLVYK-E(DEN)Q-P-W(1)DENQ;  
NAME: Ribulose-phosphate 3-epimerase family signature 1;  
CONSENSUS: LLVMAKQ-H-LLVMAFY-D-LLVMA-K-D-K(1,2)FVY-LLVMA-K-W-(ET+V);  
NAME: Ribulose-phosphate 3-epimerase family signature 2;  
CONSENSUS: LLVMAK-Q-LLVMA-K(ET)VY-K-P-K(3)Q-Q-P-W(1)H-LLVMAK;  
NAME: Arabinose 1-epimerase protein active site;  
CONSENSUS: (HS)S-T-H-H-T-PVY-H-G(L);  
NAME: Cyclophellin-type papain-like proteolytic caspase intermediate signature;  
CONSENSUS: (PT)K-D-K(ET)CMLVY-S-H-K(18)LLVMAK-LLVMA-KD-P-LLVMA-K-Q(LAQ)-G;  
NAME: Cyclophellin-type papain-like proteolytic caspase intermediate profile;  
NAME: FE3P-type papain-like proteolytic caspase intermediate signature 1;  
CONSENSUS: LLVMAK-Q-VTPY-H-KVY-L-K(1,2)H-LFT-KD-Q-K(3)D(ET)AQR-K(ET)A;  
NAME: FE3P-type papain-like proteolytic caspase intermediate signature 1;  
CONSENSUS: LLVMAKPTV-KD(KA)-K-L-K-LLVMAKPTV-KD-LLVMAK(K1)K(2)-G-K-H-LLVMAK;  
CONSENSUS: K(7)P(8)Q-K-K(2)LAQ-LFTV-G;  
NAME: FE3P-type papain-like proteolytic caspase intermediate domain profile;  
NAME: PpC-type papain-like proteolytic caspase intermediate signature;  
CONSENSUS: P-KSA-K(2)S-K-V-K(2)H-K(3)ETV-K(1)G-ETQ-K(3)K(2)S-K-Q-LLVMA;  
CONSENSUS: (GS);  
NAME: Thymoplasmin intermediate active site;  
CONSENSUS: (A)V-T-S-P-LLVMA-K-S(KA)-G-T-(GS);  
NAME: Xylose intermediate signature 1;  
CONSENSUS: LLVMAK-P-S-P-K(2)P;  
NAME: Xylose intermediate signature 2;  
CONSENSUS: P(L)H-D-G-Q-LLVMA-K(PV)-H-(DDE);  
NAME: Phosphomannose intermediate type 1 signature 1;  
CONSENSUS: Y-K-D-W-H-S-P-S;  
NAME: Phosphomannose intermediate type 1 signature 2;  
CONSENSUS: H-A-T-LLVMA-K-Q-K(2)H-LLVMA-K-S-H-A-K-S-D-H-K-LLVMA-K-A-G-T-P-E;  
NAME: Phosphogluconate intermediate signature 1;  
CONSENSUS: (DRE)K-LLVMA-K-Q-K-LFTV-S-LLVMAK-(ETV)PMAK-LLVMA-K-Q;  
NAME: Phosphogluconate intermediate signature 2;  
CONSENSUS: (DR)K-LLVMA-LLVMAKPTV-K(1)PTV-DW-K-Q-V-S-K(2)E;  
NAME: Chalconase/polychalconase 6-phosphonate intermediate signature;  
CONSENSUS: LLVMA-K(3)Q-K-LLVMA-K-LLVMA-K-Q-LLVMA-K-Q-(DPT)G-H;  
NAME: Phosphoglycerate kinase family phosphatidylcholine signature;  
CONSENSUS: LLVMA-K-S-H-G(D)K-K(3)H;  
NAME: Phosphoglucomannose and phosphomannose phosphomannose signature;  
CONSENSUS: (GSA)LLVMA-K-LLVMA-ETV(PGA)S-H-P-W(1)H(1)H(2);  
NAME: Methyltransferase-Cas intermediate signature;  
CONSENSUS: S-A-A-K-H-PTQ-KD-LLVMAKPTV-KD-(EQ)S-H-H-K(3)H-KD-D-P-K(GSA);  
CONSENSUS: G-S;  
NAME: Topoisomerase signature;  
CONSENSUS: (DQ)Q-S-W-K-Q-H-W-KSA-LLVMA-K(PV)-T-(DA);  
NAME: Ribonucleic acid epimerase 1 active site;  
CONSENSUS: (DEN)H-H-K(2)S-ETV-S-E-KD-Y-LLVMA-K(1)LLVMA;  
NAME: Ribonucleic acid epimerase 2 active site;  
CONSENSUS: (DQ)S-L-Y-(DQ)ETV-K(1,2)H-LFTV-T-S-ETV(DQ)S;  
NAME: DNA epimerase 2 signature;  
CONSENSUS: LLVMA-K-S-Q(DW)S-A-K-ETVAG;

NAME: Aminoacyl-tRNA synthetase class-I signature.  
CONSENSUS: P-AE-D-GITTAAGDEKQSLAPK+ELVMPYRRTTLLVMTACDQSDNTIQ  
CONSENSUS: ELVMPETTAADPC

NAME: Aminoacyl-tRNA synthetase class-II signature 1.  
CONSENSUS: (PTP)S+LQD+LTDLSEK+LTP+LSDDEK

NAME: Aminoacyl-tRNA synthetase class-II signature 2  
CONSENSUS: (GTAL)YF(DDEKQRLP)(GTTA)GLYMPR(DK)S(ELVMPY+ELVMTAGD)ELVMPY

NAME: WHEP-TS domain signature  
CONSENSUS: LQTVG(DHSA)+ELPY(DRL)K+LH(LKNG)(AD)H+ELPY(DKRE)  
CONSENSUS: KTHPY+KTHL+LQDLE

NAME: ATP-cases from / succinyl-CoA ligase family signature 1.  
CONSENSUS: S(ELSA)G(PST)ELVMS(GSTT)+(DQ)KTLRQ+LH(ELVMT)DA(ELVMS)G  
CONSENSUS: G-D

NAME: ATP-cases from / succinyl-CoA ligase family active site  
CONSENSUS: G-KD+A+K(7)HDTT(ELVMPY)CA(LAE)(GK)

NAME: ATP-cases from / succinyl-CoA ligase family signature 2  
CONSENSUS: G+(PT)KTH(ELVMPY)+(DA)G(SA)G(LA)(STTA)V+H+D+L(ELVMS)KTH  
CONSENSUS: G(KRE)

NAME: Chlamydomonas signature 1  
CONSENSUS: (PTW)L(D-D-S-S+LH)P(DDEKQTAR)(SA)DKK+LH(ELVMPY)

NAME: Chlamydomonas putative ATP-binding region signature  
CONSENSUS: S-P(ELVMPY)A+KTH(HPAT)G(GTAAH)G+H+K+L

NAME: Chlamydomonas class-I subfamily site  
CONSENSUS: S(ELVMS)KTH(ELVMS)D(LAE)D(LAE)D(LAE)ELVY

NAME: D-kinase-D-kinase ligase signature 1  
CONSENSUS: H-G+LH+G-S-D-G+ELVMAA(GSA)KSA

NAME: D-kinase-D-kinase ligase signature 2  
CONSENSUS: ELVY+KTH(SA)+KSAVY(S(LVCA)DELVMPKTH+LTP+LH+S  
CONSENSUS: (ELVAP)H(STP)+P(KA)

NAME: SAKCAR signature 1  
CONSENSUS: ELVMPYD(P(LVMA)S+ELVMS(ELVMA)R+LH(TA)G-S

NAME: SAKCAR signature 2  
CONSENSUS: ELVMS(ELVMA)D+L(ELVMPY)S-P-G

NAME: Polyphosphatase signature 1  
CONSENSUS: ELVMPYD+ELVMS(STTA)G-G-T(PK)G-S+(STT)KTH(ELVMS)KTH(KRE)

NAME: Polyphosphatase signature 2  
CONSENSUS: ELVMPYD(S-S+G(ELVMS)SA)G-KTH+GTT+ELVMS

NAME: Ureapote-activating enzyme signature 1  
CONSENSUS: S-A-C-S-G-S-P+P

NAME: Ureapote-activating enzyme active site  
CONSENSUS: P(ELVMS)C-T(ELVMS)KTH+(PT)P

NAME: Ureapote-activating enzyme active site  
CONSENSUS: (PTW)LDPH(LK)(PDK)ELVY+KTH+G+ELVY-C(ELVY+ELVY)

NAME: Pyruvate-carboxylase ligase signature 1  
CONSENSUS: Q(ELVMS)S-G-G+LH+G-D-Y

NAME: Pyruvate-carboxylase ligase signature 2  
CONSENSUS: V-A-T(DV)S-A-L-E+LH+G-G

NAME: Adenylsuccinate synthetase GTP-binding site  
CONSENSUS: Q+W-G-D-S-G-S-G

NAME: Adenylsuccinate synthetase active site  
CONSENSUS: G-L(KR)P+T+K+LH+K+LH

PCT/IB04/01496

1060

1061

1062



WO 81/22659

PCT/IB80/01796

CONSENSUS: [K]K[R]K[Q]K[D]K[L]YMAA[K]GSPF[P]P;  
NAME: Serum albumin family signature;  
CONSENSUS: [P]T[K]K[C]C[K]Y[C]L[P]T[K]K[L]YMPYV;  
NAME: Transferrin signature 1;  
CONSENSUS: S-K-C-T-L-W-V-E-V-L-Q-L-S-V-R-Q;  
NAME: Transferrin signature 2;  
CONSENSUS: S-F-[P]T-S-[P]T-S-T-T-A-[L]YMA-V-[T]T-F-P;  
NAME: Avidin / Streptavidin family signature;  
CONSENSUS: [D]KQ-K[D]K[L]K[L]T[A]-K[D]Y-Q-L-[D]K-[P]W-T-[K]K;  
NAME: Bacteriophage coat protein signature;  
CONSENSUS: [D]K-V-D-T-[G]A-K-[L]YMA-K-L-L-A-[L]YMAF-T-C;  
NAME: Lipocalin signature;  
CONSENSUS: [D]KQ-K-[D]KQ[K]G[K]T[A]K-K-V-L-[D]KQ[A]K-K-[L]YMPY-[C]P-Q-[C]-W-[P]T[K]KQD-K-  
CONSENSUS: [L]YMAA[K];  
NAME: Cytochrome b5p-oxidase binding protein signature;  
CONSENSUS: [Q]K-K-[Q]K-[P]T[K]W-K-[L]YMAF-K-[P]KQ-[P]T[K]KQ-K-[L]YMAF-[L]YMA-K-D;  
CONSENSUS: [L]YMAA[K];  
NAME: Aspartate-CoA binding protein signature;  
CONSENSUS: P-[T]T-A-K-[K]K-K-[L]YMAF-K-D-[L]YMAF-T-Y-[G]T-A-K-[P]T-K-Q-[T]T-A-K-D-K;  
NAME: LBP / BPI / C1PT family signature;  
CONSENSUS: [P]A-K[A]A-[L]YMAK-K-D-K-[P]Y-[T]T-K-D-L-K-D-[K]Q-K-[L]YMA-[K]KQ-  
CONSENSUS: K-D-P;  
NAME: Phosphotransferase kinase family signature;  
CONSENSUS: [P]T-K-[L]YMAF-K-D-[C]P-Q-K-P-[D]K-K-[D]K-H;  
NAME: Plant lipid transfer protein signature;  
CONSENSUS: [L]YMA-[P]A-K-D-K-K-[L]YMA-K-[L]YMA-K-[L]YMAF-T-K-[L]YMA-[T]T-K-D;  
CONSENSUS: [D]K-C-K-D-[L]YMA;  
NAME: Ubiquitin family signature 1;  
CONSENSUS: [G]A-K-D-D-K-P-K-[L]YMAF-K-D-[L]YMA-[D]E-K-[L]YMAF-D;  
NAME: Ubiquitin family signature 2;  
CONSENSUS: [D]KQ-K-[D]K-K-[D]P-K-D-[D]KQ-K-K-K-D-K-[P]KQ-[L]S-K;  
NAME: Mannosidase family signature;  
CONSENSUS: P-K-[D]E-K-[L]Y-A-T-[K]K-K-[L]KQ-[L]YMAF-T-KMAA[K]Q-V;  
NAME: Super oxidase protein signature 1;  
CONSENSUS: [L]YMAF-KQ-[L]YMAF-KQ-K-D-[L]YMAA-[D]E-K-[L]YMAF-W-A-K-D-K-K-K-K-  
CONSENSUS: [K]E-T-A;  
NAME: Super oxidase protein signature 2;  
CONSENSUS: [L]YMAF-K-D-[L]YMAF-K-D-D-K-D-K-[L]YMAF-K-D-K-[D]K-K-K-[K]K;  
NAME: Lys-T family protein/signature signature 1;  
CONSENSUS: G-[L]YMAF-K-D-[D]K-L-Q-L-[K]K-K-D-K-[L]YMAF-K-W;  
NAME: Lys-T family protein/signature signature 2;  
CONSENSUS: P-K-[L]YMAF-K-D-K-[L]YMAF-K-D-K-K-[D]T-A-[L]YMAF-K;  
NAME: PTG2 family protein/signature signature 1;  
CONSENSUS: [G]A-K[A]K-[L]YMAF-W-A-[L]YMAF-KQ-D-K-[L]YMAF-W-T-[L]YMAF-W-G-K-D-[T]A-V;  
CONSENSUS: [P]T-K-D-[D]T-A-V-K-[L]YMAF-K-D-[G]A;  
NAME: PTG2 family protein/signature signature 2;  
CONSENSUS: [P]T-K-D-[L]YMAF-W-T-[L]YMAF-W-A-K-[V]G-K-[L]YMAF-KQ-G-[G]A-K-[L]KQ;  
NAME: Aspartate-oxalate oxidase family signature;  
CONSENSUS: Y-K-D-[D]K-T-T-K-K-D-[G]T[D]K-L-K-C-K-[D]T-K-D-[L]YMAF-[L]YMAF-K-D-K-C-K-C;  
NAME: Bacterial cytochrome family signature;  
CONSENSUS: Q-Q-K-[G]A-K-D-[L]YMAF-K-W-K-W-K-[L]YMAF-K-[D]T-A-V-[L]YMAF-K-D-K;

1064

WQ 81/12659

PCT/IB06/01796

NAME: Osmot defense Osm-regulated protein signature.  
CONSENSUS: [LVMPT]wD-GwD-T + P + EwD-H[STAV]L(LVMPTV)Y.

NAME: Osmot-like domain.  
CONSENSUS: [LVMH]wD(GT) + (TA)(DA)wD(DG)(GTP)wD(LYDE)(HG)wD-  
CONSENSUS: [LQ(DS)QDS(GRDE)S.AwD(LV)wD(LVMPT)wD(LVH)wD-  
CONSENSUS: [LVMH]wD(DS) + G.

NAME: Polysaccharide oxidoreductase protein signature.  
CONSENSUS: [TH]wD-D(DPA)wD(TA)wD(TAG)(KA)(LVMPT)(DHTA)(DHT)wD-  
CONSENSUS: [STTAS]L(LVH)wD(LVPT).

NAME: Insulin-like (p)-A factor binding protein signature.  
CONSENSUS: GCG(DS)C-CwD(C-A)wD-C.

NAME: GP63/PLN30/well family signature.  
CONSENSUS: W-P(A)Y-P(LV)Q-L + (ESA)P.

NAME: CHL1/SL24 family signature.  
CONSENSUS: L + P(L)wD-T-H-H.

NAME: C1E4 protein/matrix protein signature.  
CONSENSUS: G-Q-D-Q-T-S-Q-Q-L.

NAME: Antennae signature 1.  
CONSENSUS: [PT]L(LV)G(DH)S.A-Q + (RQD)D.

NAME: Antennae signature 2.  
CONSENSUS: W-D(T)ST(A)(R) + (SR)T(DH)D(DH).

NAME: Antennae and odor-receptor protein signature.  
CONSENSUS: [LQ](LVMH)T-S(Q)APQ + (LVMPTV)H(H)PTAQwD(H)(R).

NAME: Antennae-receptor domain signature.  
CONSENSUS: [YQ](STT)wD(LVMT)wD(S)wD(DG)wD(H)wD(DT)wD(LVMT)  
CONSENSUS: wD(LVMT)wD(LVMT)wD(LVMT)wD(LVMT).

NAME: Cysteine signature.  
CONSENSUS: P-S-D-Y-A-S-P.

NAME: Chaperone light chain signature 1.  
CONSENSUS: P-L-A-Q-Q-S-S.

NAME: Chaperone light chain signature 2.  
CONSENSUS: [SR]D + S(Q)H(LVMH)(R) + (LVMH)D + L-S.

NAME: Chaperone signature 1.  
CONSENSUS: C-L-P-C-L-S + T-C.

NAME: Chaperone signature 2.  
CONSENSUS: C-L(R)wD(R) + (D)C-C(R)wD-S-C.

NAME: Chaperone signature 3.  
CONSENSUS: C-DH)T + Q-P-Q-CwD-Y-C-Y-D.

NAME: Chaperone signature 4.  
CONSENSUS: C-wD-LVH(DSDH)C(PY)(LVMH)(SA)(R)P.

NAME: Cytochrome b5 and protein "b5-like" motif signature.  
CONSENSUS: [LVMPTV]wD-L + (DSD)STT(PY)(DSD)STT)wD(PY)wD-LVH(RCTT).

NAME: Dynamic family signature.  
CONSENSUS: L-P(L)Q-G(STT)Q(D)Q(LVMH)Y-T-S.

NAME: Dynamic light chain type 1 signature.  
CONSENSUS: W-wD-G(DS) + P(ESA)S + V(DT)DHT-S.

NAME: Ph2 protein signature 1.  
CONSENSUS: H-D(T)D + Q + L-wD(LV)G + (L)AV)G(QSAP) + PwD-D.

NAME: Ph2 protein signature 2.  
CONSENSUS: [DHS]L(LVMT) + (LVMPTV)(VETAC)(STAC)G + G(QS)G-T-D(DT)G-  
CONSENSUS: [DHS]L(LVMT) + (LVMPTV)(VETAC)(STAC)G + G(QS)G-T-D(DT)G-  
CONSENSUS: [DHS]L(LVMT) + (LVMPTV)(VETAC)(STAC)G + G(QS)G-T-D(DT)G-

WQ 01/12659 PCT/IB08/01496

[illegible]

[illegible]

NAME: G-protein coupled receptors family 3 segment 1.  
CONSENSUS: (L)Y+H(L)Y(MQ)D+L-F+H(P)A(Q)H(L)Y(M)G(T)A+H(G)T(A)E(H)H(T)A(M).  
NAME: G-protein coupled receptors family 3 segment 2.  
CONSENSUS: C-C(F)Y(W)+C+L(D)C+L(H)P(Y)Y+L(L)E(W)Y+L(T)A(H)Y-C+L(D)C.  
NAME: G-protein coupled receptors family 3 segment 3.  
CONSENSUS: F+H+E(Q)T(A)E+L+L+Q(T)A(Q)F+Q(T)M.  
NAME: Vimentin segment (residue) internal binding site.  
CONSENSUS: (L)Y(M)Y+L(T)Y(M)A(C)Y(M)A(C)Y+L(T)A(L)M(E)H(Q)S(A)P(Y)Y(L)T(A)Y+L(T)H(Q)E(H)P(Y).  
CONSENSUS: (A)Y+L(T)H(Q)E(H)P(Y).  
NAME: Bacterial rhodopsin segment 1.  
CONSENSUS: S-Y+L(Q)T(Y)W+L(L)Y(M)P(Y)E(T)Y-F+L(Y)M(Q)G.  
NAME: Bacterial rhodopsin internal binding site.  
CONSENSUS: (P)T(Y)+L(P)T(Y)Q(L)Y(M)G+L(Y)M(A)P(Y)+L(T)A(L)E+L(T)H(Q)E(H)P(Y).  
NAME: Receptor tyrosine kinase class II segment.  
CONSENSUS: (D)H(L)Y(Y)Y+L(D)S-Y-T-E.  
NAME: Receptor tyrosine kinase class III segment.  
CONSENSUS: Q+H+H+H+L(T)H(M)Y+H+L+L+Q+A-C-T.  
NAME: Receptor tyrosine kinase class V segment 1.  
CONSENSUS: F+H+(D)H+L(E)A(W)Y+L(A)A-C(L)Y(M)H(S)A(L)Y(M)H(S)A(L)H(Y)H(E)H(Q)E(L)Y(A).  
CONSENSUS: H+H+L(E)H(L)C+Q(H)A(Y).  
NAME: Receptor tyrosine kinase class V segment 2.  
CONSENSUS: C+L(T)H(Q)E(H)Q(D)H(Q)W+L(D)J+P(A)Q(L)Y(M)T(H)T)+C+L+C+L+L+G+D(H)T.  
CONSENSUS: (H)S.  
NAME: Growth factor and cytokine receptors family segment 1.  
CONSENSUS: C(L)Y(Y)Y(L)Y+L(T)H(Q)E(H)P(Y)+C+L+W.  
NAME: Growth factor and cytokine receptors family segment 2.  
CONSENSUS: (T)T(S)A+L+W+D(Q)E+H+H.  
NAME: TNF(R)M(C)F(R) family cytosolic-rich region segment.  
CONSENSUS: C+H+H+P(Y)Y(W)+L+H(Q)C+L(D)J+L(C)+L(T)H(Q)E(H)P(Y)+L(T)H(Q)E(H)P(Y).  
CONSENSUS: L(D)C.  
NAME: TNF(R)M(C)F(R) family cytosolic-rich region domain.  
CONSENSUS: (P)T(Y)Y(L)E(L)H(Q)E(H)P(Y)+L+G+P+P+L+H.  
NAME: Integrin alpha chain segment.  
CONSENSUS: C+L(P)Y(W)+L(T)H(Q)E(H)P(Y)+L+G+P+P+L+H.  
NAME: Integrin beta chain segment.  
CONSENSUS: C+L(P)Y(W)+L(T)H(Q)E(H)P(Y)+L+G+P+P+L+H.  
NAME: Integrin gamma chain segment.  
CONSENSUS: C+L(P)Y(W)+L(T)H(Q)E(H)P(Y)+L+G+P+P+L+H.  
NAME: Integrin delta chain segment.  
CONSENSUS: C+L(P)Y(W)+L(T)H(Q)E(H)P(Y)+L+G+P+P+L+H.  
NAME: Integrin epsilon chain segment.  
CONSENSUS: C+L(P)Y(W)+L(T)H(Q)E(H)P(Y)+L+G+P+P+L+H.  
NAME: Integrin zeta chain segment.  
CONSENSUS: C+L(P)Y(W)+L(T)H(Q)E(H)P(Y)+L+G+P+P+L+H.  
NAME: Integrin eta chain segment.  
CONSENSUS: C+L(P)Y(W)+L(T)H(Q)E(H)P(Y)+L+G+P+P+L+H.  
NAME: Integrin theta chain segment.  
CONSENSUS: C+L(P)Y(W)+L(T)H(Q)E(H)P(Y)+L+G+P+P+L+H.  
NAME: Integrin iota chain segment.  
CONSENSUS: C+L(P)Y(W)+L(T)H(Q)E(H)P(Y)+L+G+P+P+L+H.  
NAME: Integrin kappa chain segment.  
CONSENSUS: C+L(P)Y(W)+L(T)H(Q)E(H)P(Y)+L+G+P+P+L+H.  
NAME: Integrin lambda chain segment.  
CONSENSUS: C+L(P)Y(W)+L(T)H(Q)E(H)P(Y)+L+G+P+P+L+H.  
NAME: Integrin mu chain segment.  
CONSENSUS: C+L(P)Y(W)+L(T)H(Q)E(H)P(Y)+L+G+P+P+L+H.  
NAME: Integrin nu chain segment.  
CONSENSUS: C+L(P)Y(W)+L(T)H(Q)E(H)P(Y)+L+G+P+P+L+H.  
NAME: Integrin xi chain segment.  
CONSENSUS: C+L(P)Y(W)+L(T)H(Q)E(H)P(Y)+L+G+P+P+L+H.  
NAME: Integrin omicron chain segment.  
CONSENSUS: C+L(P)Y(W)+L(T)H(Q)E(H)P(Y)+L+G+P+P+L+H.  
NAME: Integrin pi chain segment.  
CONSENSUS: C+L(P)Y(W)+L(T)H(Q)E(H)P(Y)+L+G+P+P+L+H.  
NAME: Integrin rho chain segment.  
CONSENSUS: C+L(P)Y(W)+L(T)H(Q)E(H)P(Y)+L+G+P+P+L+H.  
NAME: Integrin sigma chain segment.  
CONSENSUS: C+L(P)Y(W)+L(T)H(Q)E(H)P(Y)+L+G+P+P+L+H.  
NAME: Integrin tau chain segment.  
CONSENSUS: C+L(P)Y(W)+L(T)H(Q)E(H)P(Y)+L+G+P+P+L+H.  
NAME: Integrin upsilon chain segment.  
CONSENSUS: C+L(P)Y(W)+L(T)H(Q)E(H)P(Y)+L+G+P+P+L+H.  
NAME: Integrin phi chain segment.  
CONSENSUS: C+L(P)Y(W)+L(T)H(Q)E(H)P(Y)+L+G+P+P+L+H.  
NAME: Integrin chi chain segment.  
CONSENSUS: C+L(P)Y(W)+L(T)H(Q)E(H)P(Y)+L+G+P+P+L+H.  
NAME: Integrin psi chain segment.  
CONSENSUS: C+L(P)Y(W)+L(T)H(Q)E(H)P(Y)+L+G+P+P+L+H.  
NAME: Integrin omega chain segment.  
CONSENSUS: C+L(P)Y(W)+L(T)H(Q)E(H)P(Y)+L+G+P+P+L+H.

PCT/BO2014/01494

1069

PCT/LBDA/01596

1070





NAME: Aminoacyl tRNA synthetase signature.  
CONSENSUS: [KTS]K21N-W-K27-T-(DHS)-T.  
NAME: Biquitin ribonuclease inhibiting factor motif signature.  
CONSENSUS: [LPMAL]+[LPMPT]+[D2]+R-[SAQV]+[DTAL]+[FY]+[LKMQR]+[LPM]+[EQD]-  
K23-[LPMF].  
NAME: Channel forming surface signature.  
CONSENSUS: T-K23-W-F-P-[LPMPTF23]-K23-D.  
NAME: Multiple family cell wall protein signature.  
CONSENSUS: [LPMAG]+C-[LPMFA]-T-[LPMAG23]-K41-[LPM]-[RQ]-K23-L-[CT].  
NAME: Nucleoside diphosphate/nucleoside triphosphate kinase signature 1.  
CONSENSUS: Y-G-G-[LPT]-T-W-F-R.  
NAME: Nucleoside diphosphate/nucleoside triphosphate kinase signature 2.  
CONSENSUS: K-K23-[LPT]-K41-[LPT]-D-K23-S-K24-K25-[LPT]-Y.  
NAME: Thiol oxidoreductase signature.  
CONSENSUS: [K2]S-C-T-Q-L-W-W-W-W-(R2).  
NAME: Membrane associated cytosolic cytoplasmic / peroxide signature.  
CONSENSUS: Y-K23-[PT]-G-T-R-[PT].  
NAME: Peroxisome cytosolic (Cox1) family signature.  
CONSENSUS: P-K23-G-C-K23-[PT]-K23-C.  
NAME: Serine/threonine kinase signature.  
CONSENSUS: C-K23-[SDEKQRKSTTAA-C]-[ATDS]-[ASTD]-[ASTD]-C-[DKD]-[DEKSTTAA]-C.  
NAME: Eukaryotic protein inhibitor family signature.  
CONSENSUS: C-K23-C-K23-T-K23-C-K23-C.  
NAME: Serine/threonine kinase (Cox1) protein inhibitor family signature.  
CONSENSUS: [LPM]+D-[LPT]-[DQ]-[K23]-[LPM].  
NAME: Serine/threonine kinase signature.  
CONSENSUS: [LPMPT]-[LPMPTAC]-[DQ]-[K23]-[PT]-P-[LPMPT]-[LPMPTC]-  
[LPMFA].  
NAME: Protein inhibitor 1 family signature.  
CONSENSUS: [PTW]+[DQ]-[LPT]-G-K23-[PTW]-K23-A.  
NAME: Repeat family of serine/threonine kinase signature.  
CONSENSUS: C-P-K23-C-K23-D-K23-C-K23-C.  
NAME: Serine/threonine kinase type signature.  
CONSENSUS: C-P-K23-D-K23-P-K23-A-C-[ATD]-L.  
NAME: Cytochrome P450 signature.  
CONSENSUS: [STTQKSTV]-[LPT]-[PT]-[DQ]-[LPM]-K23-[LPT]-[LPMPTA]-  
[DQ]-[LPT].  
NAME: Threonine kinase of serine/threonine kinase signature.  
CONSENSUS: C-K23-P-K23-P-K23-C.  
NAME: Cytosolic cytoplasmic protein inhibitor family signature.  
CONSENSUS: C-K23-[AGD]-K23-[LPT]-K23-C-[R2]-[LPMPT23]-K23-G-C.  
NAME: Alpha-3-microglobulin family inhibitor signature.  
CONSENSUS: [PQ]-[SDE]-C-[PQ]-[RQ]-[LPM].  
NAME: Chaperone signature.  
CONSENSUS: C-K23-G-K23-C-[PQ]-C-K23-[PQ]-K23-C-[R2].  
NAME: Leucine-rich repeat signature.  
CONSENSUS: S-S-L-L-S-K23-[R2]-L-W-[R2]-K23-[R2]-L.  
NAME: Chaperone repeat signature.  
CONSENSUS: A-[R2]-[DQ]-K23-G-[PQ].  
NAME: Chaperone repeat signature.  
CONSENSUS: [LPMPT]-P-[R2]-[DQ]-[LPMPT]-[R2]-[RQ]-K23-[RQ]-

PCT/IB03/01496

1073

PCT/IB04/01496

1074

PCT/LB02/01496

1075

PCT/IB2009/1496

1076

PCT/L300/01496

1077

PCT/IB00/01496

1078



1079

WO 01/12679

PCT/IB98/01496

NAME: Represented end body segment 2  
CONSENSUS (LJYMYC)G-D-ESDAWQD + H-D <Z>(LBMFY)KZ(AV) <Z>GKCF <Z>  
CONSENSUS (LMPY <Z>GSAE  
NAME: RCH/EAAD MADDH family segment  
CONSENSUS (LJYMS Y + HGA)G-L-A-S-PTT) H - (LJYMS  
NAME: BLA/3WCPYAC family segment  
CONSENSUS (LJYMT AXD(LJYMPYC) PQ) T (DE)ST(A) + (PT)GA(LJYMS)G  
NAME: Uncharacterized protein family LPR020 segment  
CONSENSUS (PT)A(LPMQD)Y(G + G)LY(G + G)E  
NAME: Uncharacterized protein family LPR020 segment 1  
CONSENSUS G + V <Z>(LJY) <Z>(SA) WQ-D <Z>(LJYTY)P-H <Z>(LJYMSD-  
CONSENSUS <Z>H  
NAME: Uncharacterized protein family LPR020 segment  
CONSENSUS (LJYMS + G(LJYMT) <Z>G-C <Z>C-ET(AQ)PT)C + (LJYMS)W+G  
NAME: Uncharacterized protein family LPR020 segment  
CONSENSUS G(LJYMSD)GAH <Z>G-D <Z>(LJYMS)Q-P + L-W(D)GA(D)W+R  
CONSENSUS (LJYMSD) <Z>A <Z>T-A <Z>T-A <Z>(LJYMSD)P  
NAME: Uncharacterized protein family LPR020 segment 1  
CONSENSUS (LJYMYT)D-D-ET(A)H + H(LJYMYT)E  
NAME: Uncharacterized protein family LPR020 segment 2  
CONSENSUS P(LJYMS + LJYMS)H + H + (TA) + (DE)  
NAME: Uncharacterized protein family LPR020 segment 3  
CONSENSUS (LJYMS(LJYMS) <Z>(LJYMS)PT) <Z>(LJYMS) (LJYMS)S-T-D + P  
NAME: Uncharacterized protein family LPR020 segment  
CONSENSUS Y + G(Y)H-D-ES(A)-A-B  
NAME: Uncharacterized protein family LPR020 segment  
CONSENSUS G-D + G + P + G(Y)D(P)D-P-G  
NAME: Uncharacterized protein family LPR020 segment  
CONSENSUS (JTA) <Z>(PT)C-T-D(LJYMS) + P-P-W+G  
NAME: Uncharacterized protein family LPR020 segment  
CONSENSUS (DE)G(LJYMSD)S-T-GS-G <Z>S + G + (PT)G(LJYMSD)W+Q  
NAME: Uncharacterized protein family LPR020 segment  
CONSENSUS E(LJYMS)G-D-E-T-P(LJYMSD)A  
NAME: Uncharacterized protein family LPR020 segment  
CONSENSUS D <Z>(G)G(LJYMS)G <Z>(G)G(LJYMS) <Z>(PT) <Z>(TY) <Z>(PT)S+  
CONSENSUS (LJYMS + G(LJYMS) <Z>(PT)G  
NAME: Uncharacterized protein family LPR020 segment  
CONSENSUS L-P(Y)PT)G(LJYMS)A-T-P-A-D-A-A(LM)  
NAME: Uncharacterized protein family LPR020 segment  
CONSENSUS D-P(LJYMS)C-G-ET)G <Z>G-LB-B  
NAME: Uncharacterized protein family LPR020 segment  
CONSENSUS C-E <Z>P-W+G-E-G-LD-D-G-G-E-D  
NAME: Uncharacterized protein family LPR020 segment  
CONSENSUS D + G-LJYMS <Z>G-Y-S-G-E-E-G  
NAME: Uncharacterized protein family LPR020 segment  
CONSENSUS G + G-D(E)S + H-G-T-G + G + G-LJYMS(DC)  
NAME: Uncharacterized protein family LPR020 segment  
CONSENSUS D-Y(GJY) <Z>G-R-GT)H <Z>(LJYMS)H-P-G  
NAME: Uncharacterized protein family LPR020 segment  
CONSENSUS G(LJYMS) <Z>H + H + G(LJYMS)P + H + G + G + G <Z>(PT)G  
NAME: Uncharacterized protein family LPR020 segment

[illegible]

**We claim:**

1. An assemblage, comprising at least one nucleic acid molecule having the sequence of a clone selected from the group consisting of: hfr2\_16c16; hfr2\_16f23; hfr2\_16g18; hfr2\_16i12; hfr2\_16k22; hfr2\_16l12; hfr2\_22f21; hfr2\_22h13; hfr2\_22h13; hfr2\_22i4; hfr2\_22k3; hfr2\_22k8; hfr2\_22b10; hfr2\_22b21; hfr2\_23f2; hfr2\_23l24; hfr2\_23n16; hfr2\_23o24; hfr2\_23o5; hfr2\_2a2; hfr2\_2b17; hfr2\_2b5; hfr2\_2c1; hfr2\_2c17; hfr2\_2c18; hfr2\_2d15; hfr2\_2d17; hfr2\_2d20; hfr2\_2g18; hfr2\_2h1; hfr2\_2h10; hfr2\_2i17; hfr2\_2k14; hfr2\_2k19; hfr2\_2b16; hfr2\_2c18; hfr2\_2f16; hfr2\_2g8; hfr2\_2l2; hfr2\_41m15; hfr2\_62b11; hfr2\_62f10; hfr2\_62l19; hfr2\_62n10; hfr2\_62o17; hfr2\_64a11; hfr2\_64a15; hfr2\_64c16; hfr2\_64c4; hfr2\_64b6; hfr2\_64l20; hfr2\_64l18; hfr2\_64k24; hfr2\_64o16; hfr2\_6a17; hfr2\_6b24; hfr2\_6l20; hfr2\_6o17; hfr2\_71a20; hfr2\_72b18; hfr2\_72d13; hfr2\_72l12; hfr2\_72m16; hfr2\_72n12; hfr2\_78c24; hfr2\_78d13; hfr2\_78k24; hfr2\_78a23; hfr2\_7a24; hfr2\_7c22; hfr2\_7j4; hfr2\_82c20; hfr2\_10c20; hfr2\_82e17; hfr2\_10c17; hfr2\_82e4; hfr2\_10e4; hfr2\_82g14; hfr2\_10g14; hfr2\_82i17; hfr2\_10; hfr2\_82l24; hfr2\_10; hfr2\_82m16; hfr2\_10; hfr2\_82m6; hfr2\_10; hfr2\_1j9; hfr2\_24a15; hfr2\_24b15; hfr2\_24c23; hfr2\_24d20; hfr2\_24p5; hfr2\_3i13; hfr2\_3o17; hfr2\_46a6; hfr2\_46b10; hfr2\_46d13; hfr2\_46j20; hfr2\_46k19; hfr2\_46m4; hfr2\_47a4; hfr2\_49b; hfr2\_4c8; hfr2\_4k14; hfr2\_4m11; hfr2\_1a11; hfr2\_1c23; hfr2\_1c15; hfr2\_1g13; hfr2\_1n3; hfr2\_14g5; hfr2\_14h21; hfr2\_14p14; hfr2\_14p7; hfr2\_15a13; hfr2\_15c24; hfr2\_15c6; hfr2\_15g14; hfr2\_15h1; hfr2\_15i5; hfr2\_15j18; hfr2\_15j3; hfr2\_15k11; hfr2\_17f10; hfr2\_17f17; hfr2\_17n12; hfr2\_17n18; hfr2\_18f3; hfr2\_18f7; hfr2\_19f9; hfr2\_19f17; hfr2\_1c1; hfr2\_1g13; hfr2\_1k11; hfr2\_20c21; hfr2\_20k2; hfr2\_20m18; hfr2\_21d4; hfr2\_21j15; hfr2\_21i16; hfr2\_21n23; hfr2\_22c23; hfr2\_22g2; hfr2\_22n13; hfr2\_23i11; hfr2\_23n19; hfr2\_23o19; hfr2\_26g22; hfr2\_27d1; hfr2\_27e4; hfr2\_27o14; hfr2\_28d14; hfr2\_2a11; hfr2\_2a17; hfr2\_2d15; hfr2\_2e12; hfr2\_2f14; hfr2\_2g7; hfr2\_2h1; hfr2\_2h15; hfr2\_2i19; hfr2\_2m18; hfr2\_2m20; hfr2\_2n9; hfr2\_2o43; hfr2\_30f4; hfr2\_35b4; hfr2\_35b5; hfr2\_35c21; hfr2\_35g6; hfr2\_35k16; hfr2\_35k24; hfr2\_35n12; hfr2\_35n24; hfr2\_35n9; hfr2\_35p17; hfr2\_35p22; hfr2\_4b4; hfr2\_4f17; hfr2\_4f5; hfr2\_4h6; hfr2\_4o19; hfr2\_50j4; hfr2\_50n06;

hues3\_50a23; hues3\_6b21; hues3\_6c11; hues3\_6d16; hues3\_72k11; Hues3\_72k15;  
 hues3\_72p16; hues3\_7b22; hues3\_7d17; hues3\_7j3; hues3\_7j8; hues3\_7p10; hues3\_7p9;  
 hues3\_8e24; Hues3\_8g11; Hues3\_8g5; hues3\_8m10; Hues3\_8p7; Hues3\_9c22; Hues3\_9d20;  
 Hues3\_9k22; hute1\_17k7; hute1\_18c12; hute1\_18l19; hute1\_18u4; hute1\_18u1;  
 hute1\_19f19; hute1\_19g19; hute1\_19g22; hute1\_19h17; hute1\_19j11; hute1\_1i2;  
 hute1\_20b19; hute1\_20g21; hute1\_20h13; hute1\_20m11; hute1\_20m24; hute1\_21d15;  
 hute1\_22d2; hute1\_22e12; hute1\_22n2; hute1\_22o2; hute1\_23e13; hute1\_23g11;  
 hute1\_24c19; hute1\_24e11; hute1\_24j6; hute1\_2h3; their complements; and variants  
 thereof.

2. An assemblage, comprising at least one nucleic acid molecule having the  
 sequence of a clone selected from the group consisting of: hfr2\_16c16; hfr2\_16f21;  
 hfr2\_16g18; hfr2\_16i12; hfr2\_16k22; hfr2\_16l12; hfr2\_22f1; hfr2\_22h13;  
 hfr2\_22h13; hfr2\_22i4; hfr2\_22k3; hfr2\_22k8; hfr2\_23b10; hfr2\_23b21;  
 hfr2\_23f1; hfr2\_23i4; ; hfr2\_23n16; hfr2\_23o24; hfr2\_23o5; hfr2\_2a2;  
 hfr2\_2b17; hfr2\_2b5; hfr2\_2c1; hfr2\_2c17; hfr2\_2c18; hfr2\_2d15; hfr2\_2d17;  
 hfr2\_2d20; hfr2\_2g18; hfr2\_2h1; hfr2\_2h10; hfr2\_2i17; hfr2\_2k14; hfr2\_2k19;  
 hfr2\_2c18; hfr2\_2f16; hfr2\_2g8; hfr2\_2j2; hfr2\_4im15; hfr2\_62b11; hfr2\_62f10;  
 hfr2\_62i19; hfr2\_62n10; hfr2\_62o17; hfr2\_64a11; hfr2\_64a15; hfr2\_64c16;  
 hfr2\_64c4; hfr2\_64b6; hfr2\_64i20; hfr2\_64j18; hfr2\_64k24; hfr2\_64o16;  
 hfr2\_6a17; hfr2\_6a24; hfr2\_6a20; hfr2\_6o17; hfr2\_71o20; hfr2\_72b18;  
 hfr2\_72d13; hfr2\_72i12; hfr2\_72m16; hfr2\_72n12; hfr2\_78c24; hfr2\_78d13;  
 hfr2\_78k24; hfr2\_78n23; hfr2\_7a24; hfr2\_7a22; hfr2\_7j4; hfr2\_82c20;  
 hfr1\_10c20; hfr2\_82e17; hfr1\_10e17; hfr2\_82e4; hfr1\_10e4; hfr2\_82g14;  
 hfr1\_10g14; hfr2\_82i17; hfr1\_10; hfr2\_82i24; hfr1\_10; hfr2\_82m16; hfr1\_10;  
 hfr2\_82n6; hfr1\_10; their complements; and variants thereof.

3. An assemblage, comprising at least one nucleic acid molecule having the  
 sequence of a clone selected from the group consisting of: hfr2\_16f21; hfr2\_16k22;  
 hfr2\_22f1; hfr2\_22h13; hfr2\_22i4; hfr2\_22k3; hfr2\_22k8; hfr2\_23f1; hfr2\_23o24;  
 hfr2\_23o5; hfr2\_2a2; hfr2\_2c1; hfr2\_2c18; hfr2\_2d20; hfr2\_2g18; hfr2\_2h1;  
 hfr2\_2h10; hfr2\_2k19; hfr2\_2f16; hfr2\_2j2; hfr2\_62n10; hfr2\_64a11; hfr2\_64c16;  
 hfr2\_64c4; hfr2\_64b6; hfr2\_64i20; hfr2\_64o16; hfr2\_6a17; hfr2\_6a20; hfr2\_71o20;

hfr2\_7d13; hfr2\_72m16; hfr2\_72n12; hfr2\_78d13; hfr2\_78n23; hfr2\_7a24;  
hfr2\_7c22; hfr2\_7j4; hfr2\_82m16; and hfr1\_10.

4. An assemblage, comprising at least one nucleic acid molecule having the sequence of a clone selected from the group consisting of: hfrd2\_1j9; hfrd2\_24a15; hfrd2\_24b15; hfrd2\_24c23; hfrd2\_24n20; hfrd2\_24p5; hfrd2\_3i13; hfrd2\_3o17; hfrd2\_46a6; hfrd2\_46b10; hfrd2\_46d13; hfrd2\_46j20; hfrd2\_46k19; hfrd2\_46m4; hfrd2\_47a4; hfrd2\_466; hfrd2\_4c8; hfrd2\_4k14; hfrd2\_4m11; their complements; and variants thereof.

5. An assemblage, comprising at least one nucleic acid molecule having the sequence of a clone selected from the group consisting of: hfrd2\_1j9; hfrd2\_24c23; hfrd2\_46a6; hfrd2\_46b10; hfrd2\_46d13; hfrd2\_466; hfrd2\_4c8; their complements; and variants thereof.

6. An assemblage, comprising at least one nucleic acid molecule having the sequence of a clone selected from the group consisting of: hmcfl\_1a11; hmcfl\_1c23; hmcfl\_1c15; hmcfl\_1g13; their complements; and variants thereof.

7. An assemblage, comprising at least one nucleic acid molecule having the sequence of a clone selected from the group consisting of: hmcfl\_1c23 hmcfl\_1g13; their complements; and variants thereof.

8. An assemblage, comprising at least one nucleic acid molecule having the sequence of a clone selected from the group consisting of: hhes3\_1a3; hhes3\_14g3; hhes3\_14h21; hhes3\_14p14; hhes3\_14p7; hhes3\_15a13; Hhes3\_15c24; hhes3\_15c6; hhes3\_15g14; hhes3\_15h1; hhes3\_15i5; hhes3\_15j18; Hhes3\_15j3; hhes3\_15k11; hhes3\_17f10; hhes3\_17f17; hhes3\_17n12; hhes3\_17n18; Hhes3\_18f3; hhes3\_18f7; hhes3\_19f19; hhes3\_19j17; hhes3\_1c1; hhes3\_1g13; hhes3\_1k11; hhes3\_20c21; hhes3\_20k2; hhes3\_20m18; hhes3\_21a4; hhes3\_21j15; hhes3\_21i16; hhes3\_21n23; hhes3\_22c23; hhes3\_22g2; hhes3\_22n13; hhes3\_23i11; hhes3\_23n19; Hhes3\_23n19; hhes3\_26g22; hhes3\_27d1; hhes3\_27b4; hhes3\_27n14; hhes3\_28d14; hhes3\_2a11; hhes3\_2a17; hhes3\_2d15; hhes3\_2e12; hhes3\_2f14; hhes3\_2g7; hhes3\_2h1; hhes3\_2h15; hhes3\_2i19; hhes3\_2m18; hhes3\_2n20; hhes3\_2n9; hhes3\_2o3; hhes3\_30f4; Hhes3\_35b4; hhes3\_35h5; hhes3\_35e21;

hlex3\_35g6; hlex3\_35k16; hlex3\_35k24; hlex3\_35n12; hlex3\_35n24; hlex3\_35n69;  
 hlex3\_35p17; hlex3\_35p22; hlex3\_4b4; hlex3\_4f17; hlex3\_4f5; hlex3\_4h6; hlex3\_4o19;  
 hlex3\_50f4; hlex3\_50w06; hlex3\_50n23; hlex3\_6b21; hlex3\_6c11; hlex3\_6d16; hlex3\_72k11;  
 Hlex3\_72k15; hlex3\_72p16; hlex3\_7b22; hlex3\_7d17; hlex3\_7j3; hlex3\_7j8; hlex3\_7p10;  
 hlex3\_7p9; hlex3\_8c24; Hlex3\_8g11; Hlex3\_8g5; hlex3\_8m10; Hlex3\_8p7; Hlex3\_9c22;  
 Hlex3\_9i20; Hlex3\_9k22; their complements; and variants thereof.

9. An assembly, comprising at least one nucleic acid molecule having the sequence of a clone selected from the group consisting of: hlex3\_14g5; hlex3\_14p14; hlex3\_14p7; hlex3\_15a13; hlex3\_15g14; hlex3\_15h1; hlex3\_15j18; hlex3\_17f10; Hlex3\_18f3; hlex3\_19f19; hlex3\_19j17; hlex3\_20c21; hlex3\_21a23; hlex3\_22c23; hlex3\_22n13; Hlex3\_22n19; hlex3\_27o14; hlex3\_28d14; hlex3\_2a11; hlex3\_2d15; hlex3\_2f14; hlex3\_2g7; hlex3\_2h15; hlex3\_2i19; hlex3\_2m20; hlex3\_2a9; hlex3\_30f4; hlex3\_35gk; hlex3\_35n24; hlex3\_35p17; hlex3\_4b4; hlex3\_4f17; hlex3\_4o19; hlex3\_50f4; hlex3\_50n23; hlex3\_50w06; hlex3\_6b21; hlex3\_6d16; hlex3\_72k11; hlex3\_7d17; hlex3\_7j8; Hlex3\_8g11; Hlex3\_8g5; Hlex3\_8p7; Hlex3\_9c22; Hlex3\_9i20; Hlex3\_9k22; their complements; and variants thereof.

10. A<sub>1</sub> assembly, comprising at least one nucleic acid molecule having the sequence of a clone selected from the group consisting of: hfr2\_16g18; hfr2\_2k14; Hlex3\_35b4; hlex3\_35p22; hlex3\_7j3; hlex3\_7p10; hlex3\_20n11; their complements; and variants thereof.

11. An assembly, comprising at least one nucleic acid molecule having the sequence of a clone selected from the group consisting of: hfr2\_16c16; hfr2\_2b5; hlex3\_15i5; hlex3\_18f7; hlex3\_1k11; Hlex3\_72k15; hlex3\_7b22; hlex3\_19g22; hlex3\_24j6; their complements; and variants thereof.

12. An assembly, comprising at least one nucleic acid molecule having the sequence of a clone selected from the group consisting of: hfr2\_3d15; hlex3\_35c21; hlex3\_2b3; their complements; and variants thereof.

13. An assembly, comprising at least one nucleic acid molecule having the sequence of a clone selected from the group consisting of: hfr2\_23i24; hfr2\_21i7; hfr2\_41m15; hfr2\_62f10; hfr2\_62i19; hfr2\_64j18;

WO 91/12459

PCR/180001494

hfxd2\_24n20; hfxd2\_24p5; hfxd2\_4k14; hms3\_1g13; hms3\_21116; hms3\_23111;  
hms3\_26g22; hms3\_4b6; hms3\_72p16; hms1\_19b17; hms1\_20b13; hms1\_24e11; their  
complements; and variants thereof.

14. An assemblage, comprising at least one nucleic acid molecule having the  
sequence of a clone selected from the group consisting of: hfr2\_3g8; hfr2\_62o17;  
hfr2\_6b24; hfr2\_78b24; hfxd2\_24b15; hfxd2\_3o17; hfxd2\_66p20; hms3\_17i17;  
hms3\_17e18; hms3\_27d1; hms3\_2a17; hms3\_35b5; hms3\_35b16; hms3\_35n12;  
hms3\_35e9; hms1\_20b19; hms1\_20n24; hms1\_23e13; their complements; and variants  
thereof.

15. An assemblage, comprising at least one nucleic acid molecule having the  
sequence of a clone selected from the group consisting of: hfr2\_23b10; hfr2\_3c18;  
hfr2\_64a15; hfr2\_6o17; hfr2\_72b18; hfr2\_72i12; hfr2\_82i24(hfr1\_10);  
hms3\_14b21; Hms3\_15j3; hms3\_20n18; hms3\_22g2; hms3\_2m18; hms3\_7p9;  
hms3\_8m10; hms1\_18i1; their complements; and variants thereof.

16. An assemblage, comprising at least one nucleic acid molecule having the  
sequence of a clone selected from the group consisting of: hfr2\_23b21; hfr2\_23n16;  
hfr2\_2c17; hfr2\_62b11; hfr2\_78c24; hfr2\_82e4 (hfr1\_10e4); hfr2\_82i17  
(hfr1\_10); hfr2\_82m6 (hfr1\_10); hfxd2\_46m4; hms3\_15k11; hms3\_1c1; hms3\_1a3;  
hms3\_20k3; hms3\_21d4; hms3\_23a19; hms3\_4f5; hms3\_6c11; hms3\_8c24; hms1\_20g21;  
hms1\_22d2; hms1\_22e12; their complements; and variants thereof.

17. An assemblage, comprising at least one nucleic acid molecule having the  
sequence of a clone selected from the group consisting of: hfr2\_16i12; hfr2\_16i12;  
hfr2\_22b13; hfr2\_2b17; hfr2\_2d17; hfr2\_64d24; hfr2\_82c30 (hfr1\_10c30);  
hfr2\_82e17 (hfr1\_10e17); hfr2\_82g14 (hfr1\_10g14); hfxd2\_24a13; hfxd2\_3i13;  
hfxd2\_4m11; hmsf1\_1a11; hmsf1\_1e13; hms3\_15c6; hms3\_2a13; hms3\_27e4; hms3\_2h1;  
hms3\_33k24; hms1\_19f19; and hms1\_24c19; their complements; and variants thereof.

18. An assemblage, comprising at least one nucleic acid molecule having the  
sequence of a clone selected from the group consisting of: hfxd2\_46k19; hfxd2\_47e4;



hues3\_2c12; hues3\_21j15; hues3\_17c12; hues1\_18i19; hues1\_12; their complements; and variants thereof.

19. An assemblage, comprising at least one nucleic acid molecule having the sequence of a clone selected from the group consisting of: hues1\_17k7; hues1\_18c12; hues1\_18i19; hues1\_18d4; hues1\_18i1; hues1\_19f19; hues1\_19g19; hues1\_19g22; hues1\_19e17; hues1\_19j11; hues1\_12; hues1\_20b19; hues1\_20g21; hues1\_20b13; hues1\_20m11; hues1\_20m24; hues1\_21d13; hues1\_22c2; hues1\_22c17; hues1\_22c2; hues1\_22c2; hues1\_23c13; hues1\_23g11; hues1\_24c19; hues1\_24e11; hues1\_24g6; hues1\_2h3; their complements; and variants thereof.

20. An assemblage, comprising at least one nucleic acid molecule having the sequence of a clone selected from the group consisting of: hues1\_17k7; hues1\_18c12; hues1\_18d4; hues1\_19g19; hues1\_19j11; hues1\_22c2; hues1\_21d15; hues1\_22c2; hues1\_23g11; their complements; and variants thereof.

21. A computer readable medium, comprising in electronic form at least one nucleic acid or protein sequence of a clone selected from the group consisting of: hfr2\_16c16; hfr2\_16c21; hfr2\_16g18; hfr2\_16i12; hfr2\_16k22; hfr2\_16i12; hfr2\_22f21; hfr2\_22b13; hfr2\_22b13; hfr2\_22i4; hfr2\_22k3; hfr2\_22k8; hfr2\_23b10; hfr2\_23b21; hfr2\_23f2; hfr2\_23i24; hfr2\_23a16; hfr2\_23a24; hfr2\_23c5; hfr2\_24c2; hfr2\_2b17; hfr2\_2b3; hfr2\_2c1; hfr2\_2c17; hfr2\_2c18; hfr2\_2d15; hfr2\_2d17; hfr2\_2d20; hfr2\_2g18; hfr2\_2h1; hfr2\_2h10; hfr2\_2i17; hfr2\_2k14; hfr2\_2k19; hfr2\_3c18; hfr2\_3f16; hfr2\_3g8; hfr2\_3i2; hfr2\_41m15; hfr2\_62b11; hfr2\_62i10; hfr2\_62i19; hfr2\_62a10; hfr2\_62a17; hfr2\_64a11; hfr2\_64a15; hfr2\_64c16; hfr2\_64c4; hfr2\_64b6; hfr2\_64b20; hfr2\_64j18; hfr2\_64k24; hfr2\_64c16; hfr2\_64i7; hfr2\_64b24; hfr2\_6i20; hfr2\_6o17; hfr2\_71c20; hfr2\_72b18; hfr2\_72b13; hfr2\_72i12; hfr2\_72m16; hfr2\_72a12; hfr2\_78c24; hfr2\_78d13; hfr2\_78k24; hfr2\_78a23; hfr2\_7a24; hfr2\_7c22; hfr2\_7j4; hfr2\_82c20; hfr2\_10c20; hfr2\_82c17; hfr2\_10c17; hfr2\_82b4; hfr2\_10a4; hfr2\_82g14; hfr2\_10g14; hfr2\_82i17; hfr2\_10; hfr2\_82i24; hfr2\_10; hfr2\_82m16; hfr2\_10; hfr2\_82m6; hfr2\_10; hfr2\_10; hfr2\_24a15; hfr2\_24b15; hfr2\_24c23; hfr2\_24a20; hfr2\_24p5; hfr2\_3i13; hfr2\_3o17; hfr2\_46a6;

hfd2\_46b10; hfd2\_46d13; hfd2\_46d20; hfd2\_46d19; hfd2\_46d4; hfd2\_47a4;  
hfd2\_48b; hfd2\_4c8; hfd2\_4k14; hfd2\_4m11; hmcfl\_1a11; hmcfl\_1c23; hmcfl\_1e15;  
hmcfl\_1g13; hmc3\_1a3; hmc3\_14g5; hmc3\_14h21; hmc3\_14p14; hmc3\_14p7;  
hmc3\_15a13; hmc3\_15c24; hmc3\_15c6; hmc3\_15g14; hmc3\_15h1; hmc3\_15i5;  
hmc3\_15j18; hmc3\_15j3; hmc3\_15k11; hmc3\_17f10; hmc3\_17f17; hmc3\_17n12;  
hmc3\_17n18; hmc3\_18f5; hmc3\_18f7; hmc3\_19f19; hmc3\_19j17; hmc3\_1e1; hmc3\_1g13;  
hmc3\_1k11; hmc3\_20c21; hmc3\_20k2; hmc3\_20m18; hmc3\_21d4; hmc3\_21j15;  
hmc3\_21i16; hmc3\_21a23; hmc3\_22c23; hmc3\_22g2; hmc3\_22n13; hmc3\_23i11;  
hmc3\_23n19; hmc3\_23n19; hmc3\_26g22; hmc3\_27d1; hmc3\_27h4; hmc3\_27o14;  
hmc3\_28d14; hmc3\_2a11; hmc3\_2a17; hmc3\_2d15; hmc3\_2e12; hmc3\_2f14; hmc3\_2g7;  
hmc3\_2h1; hmc3\_2h15; hmc3\_2i9; hmc3\_2m18; hmc3\_2m20; hmc3\_2n9; hmc3\_2o3;  
hmc3\_30f4; hmc3\_35b4; hmc3\_35d5; hmc3\_35c21; hmc3\_35g6; hmc3\_35k16;  
hmc3\_35k24; hmc3\_35n12; hmc3\_35n24; hmc3\_35n9; hmc3\_35p17; hmc3\_35p22;  
hmc3\_4b4; hmc3\_4f17; hmc3\_4f5; hmc3\_4h6; hmc3\_4o19; hmc3\_50f4; hmc3\_50n06;  
hmc3\_50n23; hmc3\_6b21; hmc3\_6c11; hmc3\_6d16; hmc3\_72k11; hmc3\_72k15;  
hmc3\_72p16; hmc3\_7b22; hmc3\_7d17; hmc3\_7j3; hmc3\_7j8; hmc3\_7p10; hmc3\_7p9;  
hmc3\_8c24; hmc3\_8g11; hmc3\_8g5; hmc3\_8m10; hmc3\_8p7; hmc3\_9c22; hmc3\_9c20;  
hmc3\_9e22; hmc1\_17h7; hmc1\_18c12; hmc1\_18i19; hmc1\_18d4; hmc1\_18f1;  
hmc1\_19f19; hmc1\_19g19; hmc1\_19g22; hmc1\_19h17; hmc1\_19j11; hmc1\_1i2;  
hmc1\_20b19; hmc1\_20g21; hmc1\_20h13; hmc1\_20n11; hmc1\_20m24; hmc1\_21d15;  
hmc1\_22d2; hmc1\_22e12; hmc1\_22n2; hmc1\_22o2; hmc1\_23c13; hmc1\_23g11;  
hmc1\_24c19; hmc1\_24e11; hmc1\_24f6; hmc1\_2h3; their complements; and variants  
thereof.

22. A computer readable medium, comprising in electronic form at least one  
nucleic acid or protein sequence of a clone selected from the group consisting of:  
hfr2\_16c16; hfr2\_16f21; hfr2\_16g18; hfr2\_16i12; hfr2\_16k22; hfr2\_16l12;  
hfr2\_22f21; hfr2\_22h13; hfr2\_22h13; hfr2\_22i4; hfr2\_22k3; hfr2\_22k8;  
hfr2\_23b10; hfr2\_23b21; hfr2\_23f2; hfr2\_23f24; ; hfr2\_23o16; hfr2\_23o24;  
hfr2\_23o5; hfr2\_2a2; hfr2\_2b17; hfr2\_2b5; hfr2\_2c1; hfr2\_2c17; hfr2\_2c18;  
hfr2\_2d15; hfr2\_2d17; hfr2\_2d20; hfr2\_2g18; hfr2\_2h1; hfr2\_2h10; hfr2\_2i17;  
hfr2\_2k14; hfr2\_2k19; hfr2\_2k18; hfr2\_2f16; hfr2\_2g8; hfr2\_3j2; hfr2\_41m15;  
hfr2\_62b11; hfr2\_62f10; hfr2\_62i19; hfr2\_62n10; hfr2\_62o17; hfr2\_64a11;

hfr2\_64a15; hfr2\_64c16; hfr2\_64c4; hfr2\_64m6; hfr2\_64n20; hfr2\_64j18;  
 hfr2\_64k24; hfr2\_64o16; hfr2\_6a17; hfr2\_6b24; hfr2\_6a20; hfr2\_6o17;  
 hfr2\_71a20; hfr2\_72b18; hfr2\_72d13; hfr2\_72i12; hfr2\_72m16; hfr2\_72n12;  
 hfr2\_78c24; hfr2\_78d13; hfr2\_78k24; hfr2\_78a23; hfr2\_7a24; hfr2\_7c22;  
 hfr2\_7j4; hfr2\_82c20; hfr2\_10c20; hfr2\_82e17; hfr1\_10e17; hfr2\_82m4;  
 hfr1\_10e4; hfr2\_82g14; hfr1\_10g14; hfr2\_82i17; hfr1\_10; hfr2\_82k24; hfr1\_10;  
 hfr2\_82m16; hfr1\_10; hfr2\_82m6; hfr1\_10; complements of the nucleic acid  
 sequences; and variants thereof.

23. A computer readable medium, comprising in electronic form at least one  
 nucleic acid or protein sequence of a clone selected from the group consisting of:  
 hfr2\_16d21; hfr2\_16k22; hfr2\_22f21; hfr2\_22n13; hfr2\_22i4; hfr2\_22k3; hfr2\_22k8;  
 hfr2\_23f2; hfr2\_23e24; hfr2\_23o5; hfr2\_2a2; hfr2\_2e1; hfr2\_2e18; hfr2\_2d20;  
 hfr2\_2g18; hfr2\_2b1; hfr2\_2b10; hfr2\_2k19; hfr2\_2f16; hfr2\_2i2; hfr2\_62n10;  
 hfr2\_64a11; hfr2\_64c16; hfr2\_64a4; hfr2\_64h6; hfr2\_64i20; hfr2\_64k24;  
 hfr2\_64o16; hfr2\_6a17; hfr2\_6i20; hfr2\_71a20; hfr2\_72d13; hfr2\_72m16;  
 hfr2\_72n12; hfr2\_78d13; hfr2\_78k23; hfr2\_7a24; hfr2\_7c22; hfr2\_7j4; hfr2\_82m16;  
 hfr1\_10; complements of the nucleic acid sequences; and variants thereof.

24. A computer readable medium, comprising in electronic form at least one  
 nucleic acid or protein sequence of a clone selected from the group consisting of:  
 hfrd2\_1j9; hfrd2\_24a15; hfrd2\_24b15; hfrd2\_24e23; hfrd2\_24n20; hfrd2\_24p5;  
 hfrd2\_3i13; hfrd2\_3o17; hfrd2\_46a6; hfrd2\_46b10; hfrd2\_46d13; hfrd2\_46j30;  
 hfrd2\_46k19; hfrd2\_46m4; hfrd2\_47a4; hfrd2\_4b6; hfrd2\_4c8; hfrd2\_4k14;  
 hfrd2\_4m11; complements of the nucleic acid sequences; and variants thereof.

25. A computer readable medium, comprising in electronic form at least one  
 nucleic acid or protein sequence of a clone selected from the group consisting of: hfrd2\_1j9;  
 hfrd2\_24e23; hfrd2\_46a6; hfrd2\_46b10; hfrd2\_46d13; hfrd2\_4b6; hfrd2\_4c8;  
 complements of the nucleic acid sequences; and variants thereof.

26. A computer readable medium, comprising in electronic form at least one  
 nucleic acid or protein sequence of a clone selected from the group consisting of:

hmcfl\_1a11; hmcfl\_1c23; hmcfl\_1c15; hmcfl\_1g13; complements of the nucleic acid sequences; and variants thereof.

27. A computer readable medium, comprising in electronic form at least one nucleic acid or protein sequence of a clone selected from the group consisting of: hmcfl\_1c23; hmcfl\_1g13; complements of the nucleic acid sequences; and variants thereof.

28. A computer readable medium, comprising in electronic form at least one nucleic acid or protein sequence of a clone selected from the group consisting of: hmc3\_1a3; hmc3\_14g5; hmc3\_14b21; hmc3\_14p14; hmc3\_14p7; hmc3\_15a13; Hmc3\_15c24; hmc3\_15c6; hmc3\_15g14; hmc3\_15h1; hmc3\_15i5; hmc3\_15j18; Hmc3\_15j3; hmc3\_15k11; hmc3\_17f10; hmc3\_17f17; hmc3\_17a12; hmc3\_17a18; Hmc3\_18f; hmc3\_18f7; hmc3\_19f19; hmc3\_19j17; hmc3\_1c1; hmc3\_1g13; hmc3\_1k11; hmc3\_20c21; hmc3\_20c2; hmc3\_20m18; hmc3\_21d4; hmc3\_21j15; hmc3\_21i16; hmc3\_21a23; hmc3\_22c23; hmc3\_22g2; hmc3\_22a13; hmc3\_23i11; hmc3\_23a19; Hmc3\_23a19; hmc3\_24g22; hmc3\_27d1; hmc3\_27b4; hmc3\_27a14; hmc3\_28d14; hmc3\_2a11; hmc3\_2a17; hmc3\_2d15; hmc3\_2e12; hmc3\_2f14; hmc3\_2g7; hmc3\_2h1; hmc3\_2h15; hmc3\_2i19; hmc3\_2m18; hmc3\_2m20; hmc3\_2r9; hmc3\_2a3; hmc3\_30f4; Hmc3\_35b4; hmc3\_35b5; hmc3\_35c21; hmc3\_35g6; hmc3\_35k16; hmc3\_35k24; hmc3\_35a12; hmc3\_35c24; hmc3\_35e9; hmc3\_35p17; hmc3\_35y22; hmc3\_4b4; hmc3\_4f17; hmc3\_4f5; hmc3\_4b6; hmc3\_4e19; hmc3\_50f4; hmc3\_50w06; hmc3\_50a23; hmc3\_6b21; hmc3\_6c11; hmc3\_6d16; hmc3\_72k11; Hmc3\_72k15; hmc3\_72p16; hmc3\_7b22; hmc3\_7d17; hmc3\_7j3; hmc3\_7j8; hmc3\_7p10; hmc3\_7p9; hmc3\_8c24; Hmc3\_8g11; Hmc3\_8g5; hmc3\_8a10; Hmc3\_8p7; hmc3\_9e22; Hmc3\_9i20; Hmc3\_9c22; complements of the nucleic acid sequences; and variants thereof.

29. A computer readable medium, comprising in electronic form at least one nucleic acid or protein sequence of a clone selected from the group consisting of: hmc3\_14g5; hmc3\_14p14; hmc3\_14p7; hmc3\_15a13; hmc3\_15g14; hmc3\_15h1; hmc3\_15j18; hmc3\_17f10; hmc3\_17a18; Hmc3\_18f; hmc3\_19f19; hmc3\_19j17; hmc3\_20c21; hmc3\_21a23; hmc3\_22c23; hmc3\_22a13; Hmc3\_23a19; hmc3\_27a14; hmc3\_28d14; hmc3\_2a11; hmc3\_2d15; hmc3\_2f14; hmc3\_2g7; hmc3\_2h15; hmc3\_2i19; hmc3\_2m20; hmc3\_2r9; hmc3\_30f4; hmc3\_35g6; hmc3\_35a24; hmc3\_35p17; hmc3\_4b4; hmc3\_4f17;

hhes3\_4o19; hhes3\_5q4; hhes3\_5o23; hhes3\_5o66; hhes3\_6a21; hhes3\_6d16; hhes3\_72k11;  
 hhes3\_7d17; hhes3\_7j8; hhes3\_8g11; hhes3\_8g5; hhes3\_8p7; hhes3\_9e22; hhes3\_9i20;  
 hhes3\_9e22; complements of the nucleic acid sequences; and variants thereof.

30. A computer readable medium, comprising in electronic form at least one nucleic acid or protein sequence of a clone selected from the group consisting of: hfor2\_16g18; hfor2\_2h14; hhes3\_35p4; hhes3\_35p22; hhes3\_7j3; hhes3\_7p10; hhes1\_20m11; complements of the nucleic acid sequences; and variants thereof.

31. A computer readable medium, comprising in electronic form at least one nucleic acid or protein sequence of a clone selected from the group consisting of: hfor2\_16c16; hfor2\_2h5; hhes3\_15i5; hhes3\_18i7; hhes3\_1k11; hhes3\_72k15; hhes3\_7b22; hhes1\_19g22; hhes1\_24j6; complements of the nucleic acid sequences; and variants thereof.

32. A computer readable medium, comprising in electronic form at least one nucleic acid or protein sequence of a clone selected from the group consisting of: hfor2\_2d15; hhes3\_35c21; hhes1\_2h3; complements of the nucleic acid sequences; and variants thereof.

33. A computer readable medium, comprising in electronic form at least one nucleic acid or protein sequence of a clone selected from the group consisting of: hfor2\_23i24; hfor2\_2i17; hfor2\_41m15; hfor2\_62f10; hfor2\_62i19; hfor2\_64j18; hfor2\_24c20; hfor2\_24p5; hfor2\_4k14; hhes3\_1g13; hhes3\_21i16; hhes3\_23i11; hhes3\_26g22; hhes3\_4b6; hhes3\_72p16; hhes1\_19h17; hhes1\_20h13; hhes1\_24c11; complements of the nucleic acid sequences; and variants thereof.

34. A computer readable medium, comprising in electronic form at least one nucleic acid or protein sequence of a clone selected from the group consisting of: hfor2\_3g8; hfor2\_62o17; hfor2\_6b24; hfor2\_78k24; hfor2\_24b15; hfor2\_3o17; hfor2\_46p20; hhes3\_17i17; hhes3\_17a18; hhes3\_27d1; hhes3\_2a17; hhes3\_35b5; hhes3\_35k16; hhes3\_35m12; hhes3\_35p9; hhes1\_20h19; hhes1\_20m24; hhes1\_23e13; complements of the nucleic acid sequences; and variants thereof.

35. A computer readable medium, comprising in electronic form at least one nucleic acid or protein sequence of a clone selected from the group consisting of:

hfr2\_23b10; hfr2\_3c18; hfr2\_64a15; hfr2\_6c17; hfr2\_72b18; hfr2\_72b12;  
hfr2\_82d24(hfr1\_10); hns3\_14b11; hns3\_15j3; hns3\_20m18; hns3\_22g2; hns3\_2m18;  
hns3\_7p9; hns3\_8m10; hns1\_181; complements of the nucleic acid sequences; and  
variants thereof.

36. A computer readable medium, comprising in electronic form at least one  
nucleic acid or protein sequence of a clone selected from the group consisting of:  
hfr2\_23b21; hfr2\_23n16; hfr2\_2c17; hfr2\_62b11; hfr2\_78c24; hfr2\_82c4  
(hfr1\_10c4); hfr2\_82i17 (hfr1\_10); hfr2\_82m6 (hfr1\_10); hfr2\_46m4; hns3\_15k11;  
hns3\_1c1; hns3\_1a3; hns3\_20k2; hns3\_21d4; hns3\_23n19; hns3\_4f5; hns3\_6c11;  
hns3\_8c24; hns1\_20g21; hns1\_22d2; hns1\_22e12; complements of the nucleic acid  
sequences; and variants thereof.

37. A computer readable medium, comprising in electronic form at least one  
nucleic acid or protein sequence of a clone selected from the group consisting of:  
hfr2\_16i12; hfr2\_16i12; hfr2\_22b13; hfr2\_2b17; hfr2\_2d17; hfr2\_64d24;  
hfr2\_82c20 (hfr1\_10c20); hfr2\_82c17 (hfr1\_10c17); hfr2\_82g14 (hfr1\_10g14);  
hfr2\_24a15; hfr2\_3i13; hfr2\_4m11; hns3\_1a11; hns3\_1c15; hns3\_15c8;  
hns3\_2a3; hns3\_27k4; hns3\_2b1; hns3\_33c24; hns1\_19f19; and hns1\_24c19;  
complements of the nucleic acid sequences; and variants thereof.

38. A computer readable medium, comprising in electronic form at least one  
nucleic acid or protein sequence of a clone selected from the group consisting of:  
hfr2\_46k19; hfr2\_47a4; hns3\_2e12; hns3\_21j15; hns3\_17a12; hns1\_18i19;  
hns1\_1i2; complements of the nucleic acid sequences; and variants thereof.

39. A computer readable medium, comprising in electronic form at least one  
nucleic acid or protein sequence of a clone selected from the group consisting of:  
hns1\_17k7; hns1\_18c12; hns1\_18i19; hns1\_18d4; hns1\_18i1; hns1\_19f19;  
hns1\_19g19; hns1\_19g22; hns1\_19b17; hns1\_19j11; hns1\_1i2; hns1\_20b19;  
hns1\_20g21; hns1\_20h13; hns1\_20m11; hns1\_20m34; hns1\_21d15; hns1\_22d2;  
hns1\_22e12; hns1\_22n2; hns1\_22o2; hns1\_23e13; hns1\_23g11; hns1\_24c19;  
hns1\_24e11; hns1\_24f6; hns1\_21b3; complements of the nucleic acid sequences; and  
variants thereof.

40. A computer readable medium, comprising in electronic form at least one nucleic acid or protein sequence of a clone selected from the group consisting of: huse1\_17k7; huse1\_18i12; huse1\_18i4; huse1\_19i19; huse1\_19i11; huse1\_22n2; huse1\_21d15; huse1\_22o2; huse1\_23g11; complements of the nucleic acid sequences; and variants thereof.

41. A nucleic acid molecule having the sequence of a clone selected from the group consisting of: hfr2\_16c16; hfr2\_16f21; hfr2\_16g18; hfr2\_16i2; hfr2\_16k22; hfr2\_16i12; hfr2\_22f21; hfr2\_22h13; hfr2\_22h13; hfr2\_22i4; hfr2\_22k3; hfr2\_22k8; hfr2\_23b10; hfr2\_23b21; hfr2\_23f2; hfr2\_23i24; hfr2\_23n16; hfr2\_23o24; hfr2\_23o5; hfr2\_2a2; hfr2\_2b17; hfr2\_2b5; hfr2\_2c1; hfr2\_2c17; hfr2\_2c18; hfr2\_2d15; hfr2\_2d17; hfr2\_2d20; hfr2\_2g18; hfr2\_2h1; hfr2\_2h10; hfr2\_2i17; hfr2\_2k14; hfr2\_2k19; hfr2\_2k16; hfr2\_2c18; hfr2\_2f18; hfr2\_2g8; hfr2\_2i2; hfr2\_41m15; hfr2\_62h11; hfr2\_62f10; hfr2\_62i19; hfr2\_62n10; hfr2\_62o17; hfr2\_64a11; hfr2\_64a15; hfr2\_64c16; hfr2\_64c4; hfr2\_64b6; hfr2\_64i20; hfr2\_64j18; hfr2\_64k24; hfr2\_64o16; hfr2\_6a17; hfr2\_6b24; hfr2\_6d20; hfr2\_6o17; hfr2\_71c20; hfr2\_72b18; hfr2\_72d13; hfr2\_72i12; hfr2\_72m16; hfr2\_72n12; hfr2\_78c24; hfr2\_78d13; hfr2\_78k24; hfr2\_78n23; hfr2\_7a24; hfr2\_7e22; hfr2\_7j4; hfr2\_82c20; hfr2\_10c20; hfr2\_82e17; hfr2\_10e17; hfr2\_82e4; hfr2\_10e4; hfr2\_82g14; hfr2\_10g14; hfr2\_82i17; hfr2\_10; hfr2\_82i24; hfr2\_10; hfr2\_82m16; hfr2\_10; hfr2\_82n6; hfr2\_10; hfr2\_10; hfr2\_24a15; hfr2\_24b15; hfr2\_24c23; hfr2\_24d20; hfr2\_24p5; hfr2\_2i13; hfr2\_2j17; hfr2\_46a6; hfr2\_46b10; hfr2\_46d13; hfr2\_46i20; hfr2\_46k19; hfr2\_46m4; hfr2\_47a4; hfr2\_47b6; hfr2\_4c8; hfr2\_4h14; hfr2\_4m11; huse1\_1a11; huse1\_1c23; huse1\_1e15; huse1\_1g13; huse1\_1a3; huse1\_14p5; huse1\_14b21; huse1\_14p14; huse1\_14p7; huse1\_15a13; huse1\_15c24; huse1\_15c6; huse1\_15j14; huse1\_15h1; huse1\_15i5; huse1\_15j18; huse1\_15j3; huse1\_15k11; huse1\_17f10; huse1\_17i17; huse1\_17n12; huse1\_17n18; huse1\_18f5; huse1\_18f7; huse1\_19f19; huse1\_19j17; huse1\_1c1; huse1\_1g13; huse1\_1k11; huse1\_20c21; huse1\_20k2; huse1\_20m18; huse1\_21d4; huse1\_21j15; huse1\_21i16; huse1\_21n23; huse1\_22c23; huse1\_22g2; huse1\_22n13; huse1\_23i11; huse1\_23n19; huse1\_23o19; huse1\_23p22; huse1\_27d1; huse1\_27k4; huse1\_27o14; huse1\_28d14; huse1\_2a11; huse1\_2a17; huse1\_2d15; huse1\_2e12; huse1\_2f14; huse1\_2g7; huse1\_2h1; huse1\_2h15; huse1\_2i19; huse1\_2m18;

hnes3\_2m20; hnes3\_2m9; hnes3\_2m3; hnes3\_304; Hnes3\_35b4; hnes3\_35b5; hnes3\_35b21;  
hnes3\_35g6; hnes3\_35h16; hnes3\_35h24; hnes3\_35h12; hnes3\_35h24; hnes3\_35h9;  
hnes3\_35p17; hnes3\_35p22; hnes3\_4b4; hnes3\_4f17; hnes3\_4f5; hnes3\_4b6; hnes3\_4c19;  
hnes3\_504; hnes3\_50b06; hnes3\_50m23; hnes3\_6a21; hnes3\_6c11; hnes3\_6d16; hnes3\_72k11;  
Hnes3\_72k15; hnes3\_72p16; hnes3\_7b22; hnes3\_7d17; hnes3\_7j3; hnes3\_7j8; hnes3\_7p10;  
hnes3\_7p9; hnes3\_8c24; Hnes3\_8g11; Hnes3\_8g3; hnes3\_8m10; Hnes3\_8p7; Hnes3\_9c22;  
Hnes3\_9j20; Hnes3\_9c22; hntel\_17k7; hntel\_18c12; hntel\_18l19; hntel\_18k4; hntel\_18l1;  
hntel\_19f19; hntel\_19g19; hntel\_19g22; hntel\_19h17; hntel\_19j11; hntel\_1i2;  
hntel\_20b19; hntel\_20g21; hntel\_20h13; hntel\_20m11; hntel\_20m24; hntel\_21d15;  
hntel\_22d2; hntel\_22e12; hntel\_22a2; hntel\_22o2; hntel\_23e13; hntel\_23g11;  
hntel\_24c19; hntel\_24c11; hntel\_24f6; hntel\_2h3; their complements; and variants thereof.

42. A polypeptide encoded by the nucleic acid molecule according to claim 41.
43. An antibody or fragment thereof that is capable of binding to a specific portion of the peptide according to claim 42.
44. A pharmaceutical composition, comprising (a) an effective amount of a pharmaceutical agent, wherein said pharmaceutical agent is selected from the group consisting of the polypeptide according to claim 42, variants or functional derivatives thereof, and antibodies therein; and (2) a physiologically acceptable carrier or excipient.
45. An expression vector comprising the nucleic acid molecule of claim 41 or a fragment thereof, and optionally a promoter operably linked to said nucleic acid molecule or said fragment.
46. A method for recombinantly producing a desired peptide, comprising expressing in a host cell a peptide encoded by the nucleic acid molecule according to claim 41.